Transmission and evolution of Mycobacterium tuberculosis complex bacteria : a molecular approach

Gibson, Andrea Lucy

TRANSMISSION AND EVOLUTION OF *MYCOBACTERIUM TUBERCULOSIS* COMPLEX BACTERIA: A MOLECULAR APPROACH

Andrea Lucy Gibson

Thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

Faculty of Medicine

THE UNIVERSITY OF LONDON

King's College London

2005

ABSTRACT

Molecular techniques have greatly enhanced our understanding of the transmission and dissemination of *Mycobacterium tuberculosis* (*M.tb*), however the evolution of this organism is still poorly understood.

Molecular analysis demonstrated low levels of *M.tb* clustering (21%) in England suggesting a low rate of recent transmission. Risk factors for clustering included pulmonary smear negative disease and a past history of TB infection. A combination of Restriction Fragment Length Polymorphism (RFLP)-IS6110 and Variable Number Tandem Repeats-Mycobacterial Interspersed Repetitive Units (VNTR-MIRU) analysis provided a level of discrimination appropriate for clinical and public health action.

High levels of clustering (73%) observed in Harare, Zimbabwe, suggested a high level of recent transmission. Although no risk factors for clustering were identified, the high clustering rate may reflect the high HIV positivity in this area. A greater diversity of spoligotype strain families were observed in London compared to Harare, probably reflecting the greater international diversity of the population in London.

The most frequently seen *M. bovis* spoligotype in humans in the UK (Type 9) was also the most prevalent in UK cattle and could be subdivided by VNTR-MIRU. The majority of cases were due to reactivation of a past infection,

acquired prior to milk pasteurisation, because 72.3% of cases were over the age of 50.

A population of *M.tb* and *M. bovis* isolates previously used to construct a phylogenetic tree based on silent single nucleotide polymorphisms (sSNP), were analysed with VNTR-MIRU. VNTR-MIRU phylogenetic codes characterised 90.9% of isolates, defined *M.tb* strain families and differentiated between *M.tb* and *M. bovis*. A step-wise trend suggestive of bi-directional evolution was observed in discrepant isolates. A lineage specific panel of VNTR-MIRU could be used to sub-divide each lineage for epidemiological purposes. Single nucleotide conformation polymorphism (SSCP) followed by sequencing identified two SNPs which could be used to extend the *M.tb* complex phylogenetic tree.

DECLARATION

I Andrea Lucy Gibson declare that the work herein is all my own. My involvement in the publications listed in the front of the thesis is described as follows;

Gibson AL, Hewinson G, Goodchild T, Watt B, Story A, Inwald J, Drobniewski FA. (2004). Molecular epidemiology of disease due to *Mycobacterium bovis* in humans in the United Kingdom. J. Clin. Microbiol. Jan;42(1):431-4.

All molecular work and analysis was performed by myself (excluding deletion typing which was performed by the Veterinary Laboratory Agency). I drafted the paper and received comments and suggestions from the remaining authors.

Smith RM, Drobniewski F, **Gibson A**, Montague JD, Logan MN, Hunt D, Hewinson G, Salmon RL, O'Neill B. (2004). *Mycobacterium bovis* infection, United Kingdom. Emerg. Infect. Dis. Mar;10(3):539-41.

I performed the molecular typing and analysis for this work and was involved in the drafting of the paper.

Easterbrook PJ, **Gibson A**, Murad S, Lamprecht D, Ives N, Ferguson, A, Lowe O, Mason P, Ndudzo A, Taziwa A, Makombe R, Mbengeranwa L, Sola C, Rostogi N, Drobniewski F. (2004). High rates of clustering of strains causing tuberculosis in Harare, Zimbabwe: a molecular epidemiological study. J. Clin. Microbiol. Oct;42(10):4536-44

I performed the majority of the molecular typing, except for some of the spoligotyping which was carried out by Dianie Lamprecht. The epidemiological

analysis was performed mainly by Professor Philippa Easterbrook and Dr. Shahed Murad. I contributed to drafting the paper.

A. Gibson, T. Brown, L. Baker, F. Drobniewski. Can 15-loci MIRU-VNTR analysis provide an insight into the evolution of *Mycobacterium tuberculosis*? Appl. Environ. Microbiol. (In press)

All molecular work and analysis of data was performed by myself. I drafted the paper and received comments and suggestions on the paper from the remaining authors.

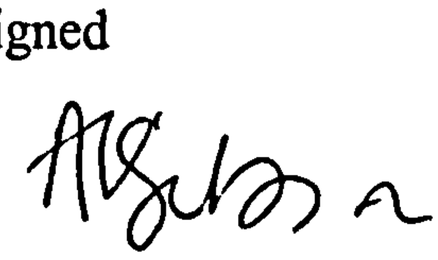
Other work within the thesis;

Chapter 3 - I performed all of the spoligotyping analysis and some of the RFLP and VNTR-MIRU. The RFLP analysis was mainly performed by Dr. Zack Fang (1998 isolates) and Dr. Krishna Gopaul (1999 isolates). The remaining VNTR-MIRU analysis was performed by Dr. Krishna Gopaul. Epidemiological data was analysed mainly by the Communicable Disease Surveillance Centre and in particular by Dr. Jane Love.

Chapter 7 – All work was performed by myself unless otherwise stated.

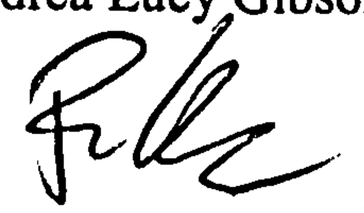
Personal communication – Some data was obtained through personal communication and has been stated within the thesis.

Signed



03/11/05

Andrea Lucy Gibson (PhD candidate)



8/11/05

Prof. Francis Drobniewski (Supervisor)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

BCG	Bacille Calmette-Guérin
BCIP	5'-Bromo-4-Chloro-3-Indolyl phosphate
bp	base pair
CAS	Central Asian family
CDSC	Communicable Disease Surveillance Centre
CI	confidence interval
CTAB	N-cetyl-NNN-trimethyl ammonium bromide
DMF	Dimethylformamide
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxy-nucleoside triphosphates
DR	Direct repeat
DVR	direct variable repeat
EAI	East African-Indian
ECL	Enhanced chemiluminesence
EDTA	ethylenediaminetetraacetic
ETR	Exact tandem repeat
HCN	high copy number
HGI	Hunter-Gaston Index
HIV	Human Immunodeficiency Virus
HPA	Health Protection Agency
ISC	Indian sub-continent
LAM	Latin American-Mediterranean
LCN	low copy number
LJ	Lowenstein-Jensen
NBT	p-Nitro Blue Tetrazolium Chloride
M. bovis	*Mycobacterium bovis*
MRU	Mycobacterium Reference Unit
MIRU	Mycobacterial interspersed repetitive units
MLST	Multilocus sequence typing
MRC	Medical Research Council
MRL	Mycobacteria Reference Laboratory

mt	mutant type
M.tb	*Mycobacterium tuberculosis*
Mycobnet	Mycobacterium resistance network
OR	Odds ratio
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase-chain reaction
PEG	polyethylene glycerol
RFLP	Restriction fragment length polymorphism
RR	rates ratio
SDS	sodium dodecyl sulphate
SSPE	saline sodium phosphate
SSCP	single stranded conformation polymorphism
sSNP	synonymous/silent single nucleotide polymorphism
nsSNP	non- synonymous single nucleotide polymorphism
TE	Tris-EDTA
TB	tuberculosis
TBE	Tris-borate-EDTA
TEMED	tetramethylethylenediamine
UK	United Kingdom
UPGMA	unweighted pair group method with arithmetic averages
VLA	Veterinary Laboratory Agency
VNTR	Variable number tandem repeats
VNTR-MIRU-12	VNTR-MIRU using 12 loci
VNTR-MIRU-15	VNTR-MIRU using 15 loci
WHO	World Health Organisation
wt	Wild-type

PUBLICATIONS

Gibson AL, Hewinson G, Goodchild T, Watt B, Story A, Inwald J, Drobniewski FA. (2004). Molecular epidemiology of disease due to *Mycobacterium bovis* in humans in the United Kingdom. J. Clin. Microbiol. 42:431-4.

Smith RM, Drobniewski F, **Gibson A**, Montague JD, Logan MN, Hunt D, Hewinson G, Salmon RL, O'Neill B. (2004). *Mycobacterium bovis* infection, United Kingdom. Emerg. Infect. Dis. 10:539-41.

Easterbrook PJ, **Gibson A**, Murad S, Lamprecht D, Ives N, Ferguson, A, Lowe O, Mason P, Ndudzo A, Taziwa A, Makombe R, Mbengeranwa L, Sola C, Rostogi N, Drobniewski F. (2004). High rates of clustering of strains causing tuberculosis in Harare, Zimbabwe: a molecular epidemiological study. J. Clin. Microbiol. 42:4536-44

A. Gibson, T. Brown, L. Baker, F. Drobniewski. Can 15-loci MIRU-VNTR analysis provide an insight into the evolution of *Mycobacterium tuberculosis*? Appl. Environ. Microbiol. Accepted 12th August 2005. Due for publication December 2005.

ACKNOWLEDGEMENTS

I am greatly indebted to all my family, friends, and colleagues whose support, advice and encouragement have made this work possible. I am immensely grateful to my two supervisors, Prof. Francis Drobniewski and Dr. Timothy Brown who have continuously supported me throughout my studies.

My special thanks to all my colleagues at the MRU for all their advice and technical support particularly the research staff; Dr. Lucy Baker, Dr. Yana Balabanova, Dr. Vladyslav Nikolayvsky and Dr. Maxine Caws for all their helpful discussion, medical and technical advice and encouragement, but above all to Dr. Krishna Gopaul for his tireless support and patience.

I would also like to thank Malcolm Yates for his technical advice and tuition and Dean Wilkinson, Poonam Gokul and Maureen Mullins for their valued friendship.

My thanks to all members of the collaborative teams involved in the study of TB in England and Zimbabwe and to the advisory committee for invaluable suggestions and advice.

My sincere gratitude to the HPA South London Virology Laboratory in particular to Dr. Melvin Smith for his advice on the SSCP work and Dr. Julian Hodgson, Kajai Jethani, and Helen Auburn for their encouragement throughout my studies.

I would like to thank King's College, London and the Health Protection Agency for the financial support making my studies possible.

A personal thanks to all my friends, particularly Louisa Hatton, who have provided me with such great support throughout my years of study.

Finally and above all, I thank my family, my Mum and Dad for their continuous love and support, Pete for his complete faith in me, Graeme and Kathy for their encouragement and my Granny and Granpa for their support and generosity throughout my studies.

CHAPTER 1.0 Introduction

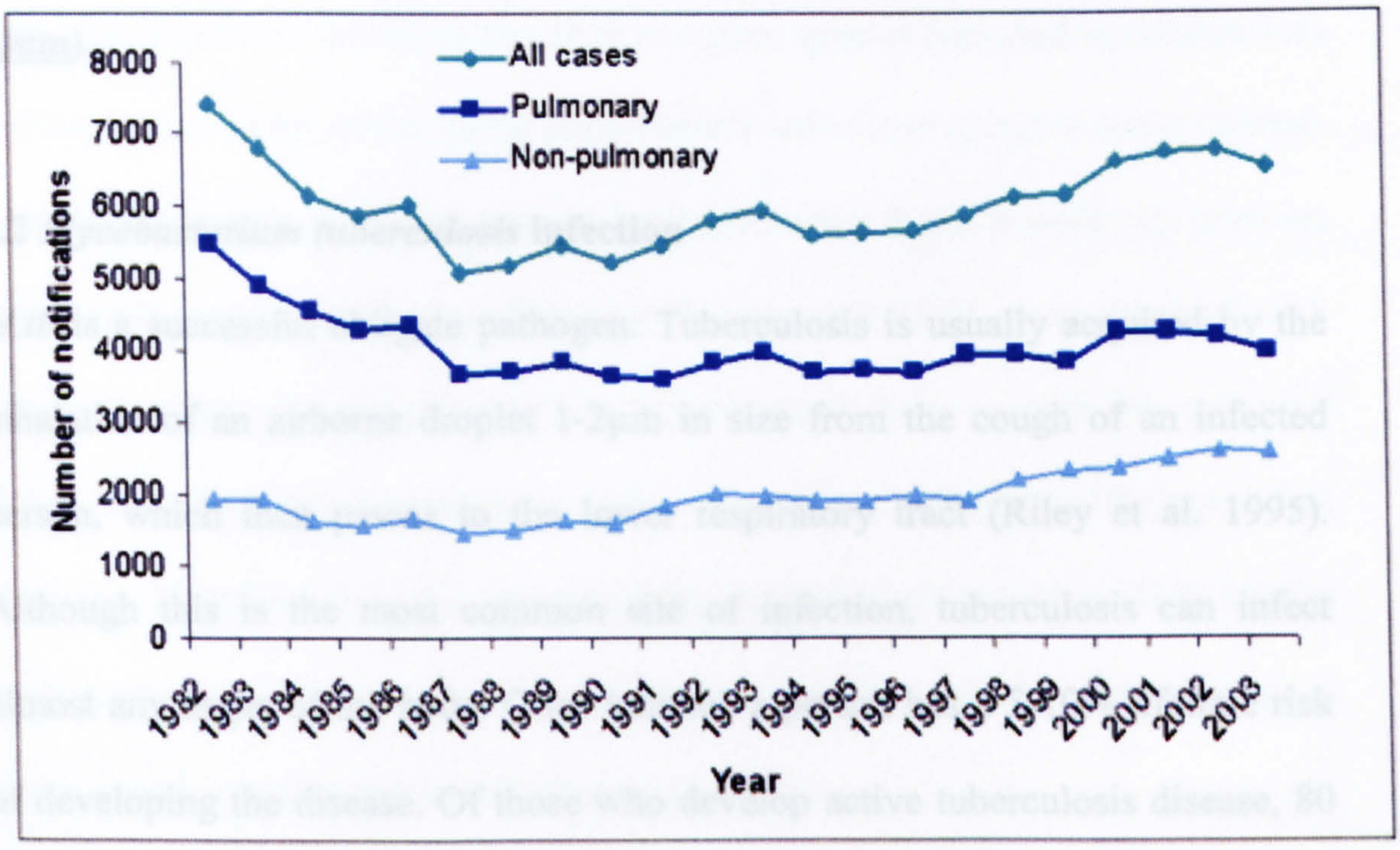
1.1 Tuberculosis

Tuberculosis (TB) is one of the greatest causes of death worldwide from an infectious disease. Approximately 2 million deaths per year are attributed to the bacillus *Mycobacterium tuberculosis* (*M.tb*), the causative agent of tuberculosis. It is estimated that one third of the world's population is infected with *M.tb* from which 8 million new cases of TB develop annually (Corbett *et al.* 2003).

The incidence of TB decreased dramatically in industrialised nations and in some parts of the developing world during the 20th century. Since the mid-1980s TB rates have increased again in the industrialised world due to increased migration of individuals from high incidence countries, poverty and a collapse of the public health infrastructure needed to control TB. In the developing world TB rates have increased even in countries which had initiated a successful TB control programme fuelled by the HIV pandemic causing the World Health Organisation (WHO) to declare TB a global emergency in 1993. Currently the global incidence of TB is increasing annually at a rate of approximately 0.4% per year, (www.int/gtb/publications/globrep/) but in sub-Saharan Africa and former Soviet Union countries the rates are in creasing much faster.

The incidence of TB in England and Wales was declining in line with many industrialised countries and dropped to a rate of between 5-6000 cases per year in the mid 1980s. Currently there are about 7000 new cases per annum (Figure 1.1).

Figure 1.1: Pulmonary and non-pulmonary tuberculosis notifications, England and Wales, 1982 - 2003



Source: Statutory Notifications of Infectious Diseases (NOIDs)

Prepared by: Communicable Disease Surveillance Centre (CDSC), Health Protection Agency Centre for Infections
http://www.hpa.org.uk/infections/topics_az/tb/epidemiology/figures/figure1.htm

In England and Wales the incidence per 100,000 population reached 12.8 in 2003, with the majority of cases in London where rates were much higher at 41.3 per 100,000 (http://www.hpa.org.uk/infections/topics_az/tb/epidemiology/table15.htm).

1.2 *Mycobacterium tuberculosis* infection

M.tb is a successful obligate pathogen. Tuberculosis is usually acquired by the inhalation of an airborne droplet 1-2μm in size from the cough of an infected person, which then passes to the lower respiratory tract (Riley et al. 1995). Although this is the most common site of infection, tuberculosis can infect almost any organ of the body. Once infected a person has a 5-10% lifetime risk of developing the disease. Of those who develop active tuberculosis disease, 80 percent will do so within two years of infection and nearly all within five years (Bloom 1994). The risk is increased in those who are HIV positive to a 10% risk per year with a 50% lifetime risk of developing the disease (www.who.int/health-services-delivery/hiv_aids/English/fact-sheet13/WHO,2000). Following infection, most people will not develop disease due to an effective immune response preventing the bacilli from proliferating leading to partial immunity to exogenous re-infection or endogenous reactivation of latent bacilli (Bloom 1994). Reactivation may occur in later life, typically in the elderly when the immune system weakens or in those who become immunocompromised for other reasons, such as illness, pregnancy or drug therapy (e.g. steroids).

1.3 Bacteriology of Mycobacteria

The genus *Mycobacterium* consists of mainly aerobic saprophytes which are closely related to Corynebacteria, Actinomyces and Nocardia. Several are clinically important including the *M.tb* complex species (detailed in section 1.4) and *M. leprae* whilst others cause opportunistic infections or never cause disease. Most mycobacteria are slow growers taking between 2 and 8 weeks to grow on solid media; typically the doubling time for example of *M. tb* is about 17 to18 hours (Ratledge and Stanford 1982). Mycobacteria have a lipid-rich cell wall which is impermeable to acidified organic compounds giving them their acid-fast property. Upon staining, for example with Zeihl-Neeson staining, *M.tb* can be seen under a microscope as a rod shaped bacillus which often display a characteristic cording pattern (Collins *et al.* 1997).

1.4 *Mycobacterium tuberculosis* complex bacteria

M.tb, M. bovis, the vaccine strain BCG, *M. africanum, M. microti,* and *M. canetti* form the *Mycobacterium tuberculosis* complex, with *M.tb* being the most frequently isolated species in cases of human tuberculosis. *M. bovis* accounts for 1% of human cases in the UK (http://www.hpa.org.uk/infections/topics_az/tb/Mbovis), while *M. africanum* is more often seen in patients from Africa (Niemann *et al.* 2002). *M microti* rarely causes disease in humans but commonly causes disease in other mammals (Cavanagh *et al.* 2002, Oevermann *et al.* 2004) and *M. canetti* appears to be a rare cause of disease in humans (van Soolingen *et al.* 1997). The *M.tb* complex bacteria are differentiated on their phenotypic and biochemical characteristics (Table 1.1) however traditional bacteriology is time

consuming therefore there has been a move towards molecular techniques for the identification of these species.

Table 1.1 Principle characteristics of the species within *M.tb* complex (Based on Collins *et al.*1997)

	Characteristic					
	Pyrazinamide[a]	O_2[b]	TCH[c]	Nitrate reduction[d]	Pyruvate-Glycerol[e]	Colonies[g]
M.tb	S	A	+	+	P=G[f]	E
M. africanum II	S	M	-	+	P=G	E
M. africanum I	S	M	-	-	P	D
M. microti	S	M	-	-	P	D
M. bovis	R	M	-	-	P	D
M. bovis BCG	R	A	-	+/-	G	E

[a] R - resistant / S- sensitive. [b] A - aerobic growth / M - microaerophilic growth
[c] Growth in the presence of TCH (thiophen-2-carboxylic acid hydrazide)
[d] Nitrate reduced to nitrite [e] Preferred growth substrate. [f]P=G – Grows well on both types of media. [g]E – Eugonic / D – Disgonic

1.5 Molecular analysis of *Mycobacterium tuberculosis* complex bacteria

Advances in molecular typing techniques over recent years have greatly enhanced our understanding of the transmission and dissemination of *M.tb*. Molecular tools are invaluable in outbreak investigations as they enable the rapid identification of clusters of related cases that may not be linked by traditional epidemiological methods and conversely separate cases that appear to be epidemiologically linked but are not (van Soolingen 2001, Drobniewski *et al.*

2003). In turn, this has increased the speed at which outbreaks are identified and furthermore permits a more efficient surveillance system for the control of tuberculosis.

In epidemiological terms the organisms involved in an outbreak are clonally related. A cluster of cases according to molecular data is defined as two or more isolates with an indistinguishable fingerprint. A cluster in a population is thought to imply recent transmission whereas unique fingerprints are thought to be epidemiologically unrelated and so have arisen from the reactivation of a latent infection or the immigration of patients who have acquired infection from abroad (Alland *et al.* 1994, Maguire *et al.* 2002). It is essential to identify and differentiate between the levels of recent transmission versus reactivation in a population as this greatly affects the control program required. A country with high rates of recent transmission would need to target resources at contact tracing and screening which would be less important in a population with high rates of reactivation.

Many molecular typing tools are available for the analysis of *M.tb* (van Soolingen 2001, Mostrom *et al.* 2002). Restriction fragment length polymorphism (RFLP), based on the insertion element IS*6110*, is considered to provide the best discrimination of *M.tb* isolates and is regarded as the gold standard for strain typing *M.tb* (van Embden *et al.* 1993, Kremer *et al.* 1999). However, the discriminatory power of RFLP-IS*6110* is reduced in low copy number (LCN) isolates i.e. those with less than 5 copies of IS*6110*. Secondary typing techniques such as spacer oligonucleotide typing (spoligotyping) and

Variable Number of Tandem Repeats (VNTR) analysis can be used to differentiate between LCN isolates, although there are significant differences in their discriminatory power (Filliol *et al.* 2000).

Although RFLP-IS*6110* is considered the gold standard it is a slow, cumbersome technique which requires viable cultures. The development of rapid PCR techniques in an effort to replace RFLP-IS*6110*, until recently, have not provided the level of discrimination needed to accurately define clusters. Most have tended to link additional unrelated cases to clusters, which in terms of an outbreak investigation would lead to increased and potentially costly contact tracing of the linked cases. A promising new technique, analysis of Mycobacterial Interspersed Repetitive Units (MIRU) described by Supply *et al.* (2001), has been developed. MIRU is a rapid PCR-based technique, and is a form of VNTR analysis as originally devised by Frothingham and Meeker-O'Connell (1998) whereby typing is based on the number of repetitive units found at certain loci within the genome. Kwara *et al.* (2003) demonstrated that a combination of RFLP-IS*6110* with MIRU as a secondary technique provided better discrimination than the commonly used RFLP-IS*6110* plus spoligotyping. It has been argued that MIRU should in fact be used as a first line technique, and then followed up with RFLP-IS*6110* to define clusters (Savine *et al.* 2002). Other studies have cited MIRU as a possible replacement for RFLP-IS*6110* and have shown similar levels of discrimination (Supply *et al.* 2001, Hawkey *et al.* 2003) although some studies have refuted this claim (Nguyen *et al.* 2004, Gopaul *et al.* unpublished data). As a PCR-based technique, MIRU is more rapid then RFLP. MIRU produces a numerical code rather then a visual banding pattern which is open to

interpretation that can be subjective. A numerical code has the advantage of being truly portable and permits straightforward analysis and comparison on a global scale.

VNTR typing has been applied to many organisms including man where it is the basis of forensic fingerprinting and for deciphering familial relationships. These repetitive units or minisatellites are defined as >8bp in length and as yet their function is unknown. Microsatellites are smaller in length (<9bp) and allelic variations in these repetitive units have been linked to mental health problems, neurological disorders and cancer in humans (Reddy *et al.* 1999, Mandel and Biancalana 2004, Claij and te Riele 1999). In pathogenic bacteria the study of tandem repeats have shown they may have a gene regulatory function (Supply *et al.* 1997) or an effect on virulence (Moxon *et al.* 1994, van Belkum *et al.* 1997). Supply *et al.* (2000) identified 41 minisatellites in *M.tb* H37Rv which were named MIRU and ranged from 40 to 100bp. Of these 41 loci, 12 which displayed variation in copy number were used initially in the strain typing scheme as described above. The analysis of the complete genome sequence of the laboratory strain *M.tb* H37Rv has shown there are 79 of these repeat regions (http://minisatellites.U.psud.fr, based on a unit length of between 0 and 2000 bases with no overlapping repeats) therefore additional hypervariable regions exist that have not been examined. These minisatellites were identified in *M.tb* H37Rv therefore it is conceivable that other strains of *M.tb* will contain additional variable regions that could potentially be used to improve the discriminatory power of MIRU.

1.6 Evolution of *Mycobacterium tuberculosis*

The discovery of *M. tuberculosis* DNA in the lungs of mummies from both pre-Columbian America and ancient Egypt dating back to 1000-1300 AD and 1550-1080 BC respectively, suggests that *M.tb* has caused disease in humans worldwide for thousands of years (Salo *et al.*1994, Nerlich *et al.* 1997). Despite this, relatively little is known about the evolution of this organism. Until recently *M.tb* was believed to have evolved from *M. bovis* following the domestication of cattle partly because of its small host range compared to the broad mammalian host range of *M. bovis*. However the complete genome sequence of both organisms (Cole *et al.* 1998, Garnier *et al.* 2003) has shown this is unlikely to be true. The genome of *M. bovis* is much smaller than that of *M.tb* but has more than 99.95% identity and contains no unique genes suggesting that *M. bovis* may have originated from *M.tb*.

Molecular techniques, such as deletion typing, have allowed the classification of *M.tb* complex strains into phylogenetic lineages (Brosch *et al.* 2002). This technique is based on the presence or absence of regions of difference in the genome. The absence of one of these regions (region of difference 9) from *M. bovis* which is present in *M.tb* and *M. canetti* supports the evolution of *M. bovis* from *M.tb*. Recently additional deleted regions have been identified in *M. bovis* using an Affymetrix GeneChip system which has lead to further differentiation of the *M. bovis* lineage (Mostowy *et al.* 2005).

Numerous studies have used molecular methods to elucidate the phylogeny of *M.tb* (Fomukong *et al.* 1997, Fang *et al.* 1998, Fang *et al.* 1999a, Dale *et al.* 2003, Brosch *et al.* 2002, Filliol *et al.* 2003, Sola *et al.* 2001a, Sola *et al.* 2001b, Baker *et al.* 2004). The insertion element IS*6110*, the gold standard for strain typing, has also been used for evolutionary studies. It has been shown that IS*6110* types can be grouped into superfamilies (Dale *et al.* 2005) and that isolates with low copy numbers of IS*6110* form families of strains which are unlikely to have arisen through convergent evolution by the insertion of the IS*6110* element into preferred positions (Dale *et al.* 2003). The IS*6110* element is commonly found in the direct repeat (DR) region on which spoligotyping is based. Strains lacking this element were also found to lack complete direct variable repeats (DVR) suggesting IS*6110* movement, facilitated by homologous recombination, may remove these DVRs (Fang *et al.* 1998). The evolution of the DR locus is thought to be unidirectional, i.e. spacers are lost over time through both the transposition of the IS*6110* element and through deletion of single or stretches of contiguous DVRs by homologous recombination between neighbouring or distant DRs (van Embden *et al.* 2000).

Genotyping techniques, such as multilocus sequence typing (MLST), have been successfully used to characterise several bacterial populations based on neutral genetic variation at multiple sites in the genome (Maiden *et al.* 1998, Enright *et al.* 1998, Dingle *et al.* 2001, Gutacker *et al.* 2002, Baker *et al.* 2004). Sreevatsan *et al.* (1997) demonstrated that two non-synonymous polymorphisms, one in catalase-peroxidase (*kat*G) and one in the A subunit of gyrase (*gyr*A), could divide *M.tb* into 3 genetic groups. Group 1 contained the *M.tb* ancestral strains.

MLST uses multiple synonymous mutations that are not subjected to evolutionary selection i.e. are neutral, making them valuable for studying evolutionary relationships. The *M.tb* genome is thought to be highly conserved as sequencing of structural genes and host immune system protein targets has shown limited genetic variation (Sreevatsan *et al.* 1997; Musser *et al.* 2000). A lack of variation could hinder MLST, however Baker *et al.* (2004) demonstrated that neutral variation could be used to index *M.tb* into 4 lineages with *M.bovis* clearly separated on a different branch of the tree.

Identifying mutations in the genome is commonly performed using DNA sequencing. DNA sequencing is a relatively expensive technique therefore other methods, which are more rapid and cheaper, can be used to screen for mutations. These include, single-stranded conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), heteroduplex and mix-match cleavage. The main disadvantage of these methods is that although a mutation can be easily identified it is not characterised as with DNA sequencing.

SSCP has been successfully used to screen for mutations (Pretorius *et al.* 1995, Fang *et al.* 1999b, Cardoso *et al.* 2004) by identifying conformational changes in the single-stranded DNA of mutant strains. These alternations cause a change in mobility on polyacrylamide gels when compared to wild-type strains resulting in a shift of band position. Although the sensitivity of this technique varies (Telenti *et al.* 1997, Bobadilla-del-Valle *et al.* 2001, Cardoso *et al.* 2004) SSCP has been successfully used to identify polymorphisms that can be characterised with direct sequencing.

1.7 Aims and Objectives

1) Examine the transmission of *M.tb* and *M. bovis*

   a. Assess the distribution and transmission rates of *M.tb* in a low incidence country, England, using a combination of molecular typing techniques and epidemiological information. To examine the occurrence of clusters and identify any risk factors for clustering. Determine the role of molecular typing as a tool for the identification of related clusters of TB for which clinical and public health action might be appropriate and for national surveillance.

   b. Investigate the pattern of TB transmission in a high incidence area Harare, Zimbabwe using several molecular techniques and epidemiological data. To examine the role of recent transmission versus reactivation. Identify risk factors for clustering and determine the impact of HIV in a country where HIV is endemic.

   c. Compare and contrast any spoligotyping families present in Harare, Zimbabwe (a high incidence population) to those seen in London, UK (a low incidence population).

   d. Examine the molecular epidemiology of *M. bovis* infection in humans in the UK using molecular typing techniques and compare the typing patterns observed to those seen in UK cattle.

2) Investigate the evolution of *M.tb complex* strains

   a. Examine the phylogeny of *M.tb* complex strains by comparing VNTR-MIRU profiles to lineages defined by synonymous single

nucleotide polymorphisms (sSNP). Determine if the phylogeny of these organisms can be defined using VNTR-MIRU as a rapid tool which could aid public health planning.

b. Expand our understanding of the phylogeny of the *M. bovis* by identifying potential sSNP which could be used to extend a previously published phylogenetic tree (Baker *et al.* 2004).

CHAPTER 2.0 - Materials and Methods

2.1 Bacterial culture

All isolates were cultured onto one slope of Lowenstein-Jensen (LJ) media containing pyruvate. All slopes were incubated at 37°C for 8 weeks. Stock cultures were stored at -70 °C in Middlebrook 7H9 media with 10% OADC in plastic cyrovials.

2.2 DNA preparation

2.2.1 DNA extraction - method 1

A crude extraction method was used when only PCR methods were required. A loop of culture was removed using a 1µl loop and placed in 150 µl of sterile dH_2O. An equal volume (150 µl) of chloroform was added. The mixture was vortexed for 10 s and then heat treated at 80°C for 20 mins to kill the bacteria. The extracted culture was frozen at –20°C for 10 mins or until needed. After, thawing, the mixture was spun at 8000 x g for 5 mins before use.

2.2.2 DNA extraction - method 2

Purified genomic DNA for RFLP-IS*6110* analysis was extracted as follows: approximately half the bacterial growth on a LJ slope was removed with a sterile swab, re-suspended in 600µl of TE buffer (10mM Tris pH8, 1mM EDTA) and heat treated at 80°C for 20 mins to kill the bacteria. Tubes were allowed to cool before adding 50µl lysozyme (10mg/ml) and vortexed. The mixture was then incubated overnight at 37°C. After incubation, 10µl proteinase K (10mg/ml) and

35µl sodium dodecyl sulfate (SDS) 20% (w/v) was added, vortexed and incubated at 55°C for 30 mins. 100µl of 5M NaCl was added and vortexed before adding 100µl of prewarmed N-cetyl-NNN-trimethyl ammonium bromide (CTAB) solution. The mixture was vortexed again and incubated at 65°C for 15 mins. The mixture was allowed to cool to room temperature before adding 700µl chloroform:isoamyl alcohol (24:1). The mixture was then vortexed for 10 sec and spun at 8000 x g for 8 mins. The supernatant was pipetted into a separate clean tube and 600µl of cold (-20°C) isopropanal-2-ol was added to precipitate the DNA. The mixture was inverted several times and then incubated for 1 hr at -20°C. After incubation the mixture was spun at 8000 x g at room temperature for 15 mins. The supernatant was decanted off being careful not to lose the DNA pellet from the tube. The pellet was washed with 1 ml of 70% ethanol and the tube was inverted to mix then spun at 8000 x g for 5 mins before removing the liquid with a pipette. The sample was re-spun for 1 min and any remaining alcohol removed before being left to air-dry. The pellet was re-suspended in 30µl TE and left to dissolve at 4°C overnight.

2.3 Restriction fragment length polymorphism (RFLP) analysis

RFLP-IS*6110* analysis was performed according to the recommended standardised protocol described by van Emben *et al.* 1993.

2.3.1 Transfer of digested genomic DNA fragments.

For each sample, 4.5 µg genomic DNA was digested for two hours with *Pvu*II and run out on a 0.8% agarose gel for approximately 16 hours at 30V. An external marker of 1 µg *Pvu*II digested MT14323 DNA was run in lanes 4, 13

and 27 on of a 30 lane gel to enable standardisation between blots. In addition 1 µg of lambda/*Hind*III was run as a size standard. Each gel was run until the 2.0kb lambda/*Hind*III fragment had migrated 7cm from the well when viewed under UV light. The gel was then treated with 0.25M hydrochloric acid for 20 minutes to depurinate the DNA fragments, rinsed in distilled water and treated twice for 15 mins in denaturing solution (0.5M NaOH, 1.5M NaCl). Following a final rinse in distilled water, the gel was treated for 30 mins in neutralising solution (1M Tris pH 7.4, 1.5M NaCl).

Transferring the digested DNA fragments from gel to nitrocellulose membrane involved the use of a vacuum blotter (Hybaid, Middlesex, UK). Briefly a sheet of 2X SSC (0.3M NaCl and 0.3M Sodium Citrate) soaked 3MM filter paper measuring 25 x 20 cm was placed on the base of the blotter and rolled with a 10 ml pipette to remove any bubbles formed. A 15 x 20 cm sheet of Hybond-N nylon membrane (Amersham International plc, Buckingham, UK), moistened with 2X SSC was then layered carefully on top and again rolled to remove any bubbles present. A rubber mask with a central portion of 14 X 19cm missing was placed over the top. The gel was then aligned over the exposed membrane. The vacuum was applied for 90 mins to transfer the DNA to the membrane. The membrane was rinsed in 2X SSC and UV crosslinked with a Stratalinker 1800 (Stratagene, La Jolla, California) using a setting of 12, 000 µJ/cm^2 for 2 mins.

2.3.2 Preparation and labelling of IS*6110* probe

The IS*6110*-probe was generated by performing DNA amplification with the following primers; INS1 – 5' CGT GAG GGC ATC GAG GTG GC 3' and INS2

– 5' GCG TAG GCG TCG GTG ACA AA 3'. The reaction mix (total volume 50µl) contained; 1x NH_4 Buffer (Bioline, London, UK), 1.5mM $MgCl_2$, 0.2mM sterile dNTPs, 0.4µM INS1, 0.4µM INS2, 10ng *M. bovis* BCG (Glaxo, Middlesex, UK) genomic DNA and 1U Taq polymerase (Bioline, London, UK), with the remaining volume made up by dH_2O. Amplification was carried out in a 0.2µl thin-walled PCR tube (Alpha, Eastleigh, UK) in a Perkin Elmer 9600 Thermocycler (Perkin Elmer Cetus, USA) for 5 mins at 94°C then 30 cycles of; 30 s at 94°C, 30 s at 58°C, 30 s at 72°C and one cycle of 5 mins at 72°C. The resulting amplified mixture was then run and excised from a 0.8% agarose gel, melted and dissolved in an equal volume of sterile dH_2O to give a final concentration of 0.2%. The probe was labelled with Biotin using ECL Direct™ Nucleic Acid Labelling and Detection System (Amersham International plc, Buckingham, UK) according to the manufacturer's instructions.

Hybridisation and detection of the IS*6110* probe was performed using the same ECL Direct™ Nucleic Acid Labelling and Detection System kit (Amersham International plc, Buckingham, UK).

2.4 Spoligotyping

Spoligotyping was performed using the standardised protocol described by Kamerbeek *et al.*, 1997.

2.4.1 PCR Analysis of the Direct Repeat (DR) region

The reaction mix (total volume 20µl) contained 16mM NH_4SO_4 reaction buffer (Bioline, London, UK), 200µM dCTP, dGTP, dATP and dTTP (Pharmacia Biotech, UK), 5% (v/v) Dimethylsulphoxide (DMSO), 10pmol Dra primer (5'–GGT TTT GGG TCT GAC GAC-3' (biotinylated at the 5' end)) and DRb primer (5'-CCG AGA GGG GAC GGA AAC-3') (Isogen Bioscience, Maarssen, The Netherlands), 1U Taq polymerase (Bioline, London, UK) and 1µl template DNA. Amplification was carried out in a 0.2µl thin-walled PCR tube (Alpha, Eastleigh, UK) in a Perkin Elmer 9600 Thermocycler (Perkin Elmer Cetus, USA) for 2 mins at 93°C then 30 cycles of; 30 s at 93°C, 1 min at 55°C, 1 min at 72°C and one cycle of 10 mins at 72°C. Two positive controls (*Mycobacterium tuberculosis* H37Rv and *Mycobacterium bovis BCG* P3) (10 ng/µl of genomic DNA) and 2 negative controls (sterile dH_2O) were included with each run.

2.4.2 Hybridisation

15µl of amplified product was added to 150µl of 2x saline sodium phosphate (SSPE) (0.2M phosphate buffer, pH 7.4 containing 2.98M NaCl and 0.02M EDTA) /0.1% SDS and denatured by heating at 100°C for 10 mins and immediately put on ice. A BiodyneC membrane (Isogen Bioscience, Maarssen, The Netherlands) with the 43 spacers covalently attached (Appendix 1) was washed in 250ml of 2x SSPE/0.1% SDS for 5 mins at 60°C. The membrane and a support cushion were placed in the miniblotter MN45 (Immunetics, Cambridge, Mass, USA) and the residual fluid removed by aspiration. 150µl of each PCR product was loaded in a separate lane on the miniblotter and allowed to hybridise at 60°C for 90 min. The fluid was aspirated from each lane before the miniblotter

was opened. The membrane was washed twice with 250ml of 2x SSPE/0.5% SDS in a hybridisation bottle for 10 mins at 60°C and allowed to cool before adding the enzyme to prevent inactivation. The membrane was washed with 5µl of 500U/ml streptavidin-peroxidase conjugate (Roche, East Sussex, UK) in 20ml of 2x SSPE/0.5% SDS for 60 mins at 42°C and then washed twice in 250ml of 2x SSPE/0.5% SDS at 42°C for 10 mins. The membrane was rinsed twice in 2x SSPE for 5 mins at room temperature.

2.4.3 Detection

The hybridised DNA was detected using an ECL kit (Amersham International plc, Buckingham, UK). The membrane was incubated in 20 ml of detection fluid for 1 min and then placed in a plastic bag and exposed to autoradiographic film for 20 min. The film was developed and fixed using standard photographic grade reagents (Sigma-Aldrich, Poole, UK).

2.5 Variable Number Tandem Repeats (VNTR)

VNTR typing was performed using various combinations of tandem repeat loci. The nomenclature used throughout the thesis for each combination of tandem repeats is described in table 2.5.1.

Table 2.5.1 VNTR nomenclature

Nomenclature	Loci utilised	Identified by;
VNTR-ETR	ETR A to E	Frothingham and Meeker-O'Connell 1998
VNTR-MIRU-12	12 loci MIRU	Supply *et al.* 2001
VNTR-MIRU-15*	12 loci MIRU plus ETR A to C	As above

* The two remaining ETRs D and E are included within the VNTR-MIRU-12 and correspond to MIRU 4 and 31

2.5.1 VNTR-Exact Tandem Repeats (VNTR-ETR)

When only the analysis of VNTR-ETR loci was required, the protocol described by Frothingham and Meeker-O'Connell (1998) was performed. Briefly, each tandem repeat was amplified in a single PCR reaction and subjected to agarose gel electrophoresis with 100bp and 20bp size markers. The size of each PCR product corresponds to the number of tandem repeats at each locus. The original ETR-B forward and reverse primers were extended by 3 bases at the 3' end (Appendix 2) to increase the annealing temperature for use with Hotstar Taq (Qiagen, West Sussex, UK).

2.5.2 VNTR-Mycobacterial interspersed repetitive units (VNTR-MIRU)

The method and primers for VNTR-MIRU-12 were as described by Kwara *et al.*, (2003). VNTR-MIRU-15 was performed using the same method. (For both VNTR-MIRU-12 and VNTR-MIRU-15 primers see Appendix 2).

2.5.2.1. PCR analysis of tandem repeat regions

Multiplex PCR reactions for VNTR-MIRU-12 were set up as follows; the reaction mix (total volume 10µl) contained 2.5µl sterile dH_2O, 5µl 2 x reaction buffer (3mM $MgCl_2$, 3.2mM dNTPs, 2x NH_4 buffer), 1.25µl of each primer mix (2µM), 0.1µl Taq polymerase (5U/µl) (Bioline, London, UK) and 1µl template DNA. A primer mix comprised of forward and reverse primers for one locus. Forward primers were labelled with one of the three standard Beckman Coulter dyes (D2, D3, or D4). For each multiplex reaction two primer mixes were used corresponding to each of the loci analysed. The loci analysed in each multiplex reaction are shown in table 2.5.2. Both multiplex and simplex PCR reactions were set up for VNTR-MIRU-15 and are shown in Table 2.5.3. Where only one locus was analysed, 2.5µl of primer mix was added.

Table 2.5.2 PCR reactions for VNTR-MIRU-12

PCR reaction	PCR 1[a]	PCR 2	PCR 3	PCR 4	PCR 5	PCR 6
VNTR-MIRU loci analysed	4	20	2	31	10	26
VNTR-MIRU loci analysed	16	39	24	23	27	27
Dye[b]	D2	D2	D3	D3	D4	D4

[a] Multiplex PCR reactions were performed containing two different primer sets corresponding to two different loci

[b] PCR primers were labelled with standard Beckman Coulter Dyes (D2, D3 and D4)

Table 2.5.3 PCR reactions for VNTR-MIRU-15

PCR reaction[a]	PCR 1	PCR 2	PCR 3	PCR 4	PCR 5	PCR 6	PCR 7	PCR 8	PCR 9
VNTR-MIRU loci analysed	4	39	20	2	31	C	10	27	26
VNTR-MIRU loci analysed	16	A		24	40		23	B	
Dye[b]	D2	D2	D2	D3	D3	D3	D4	D4	D4

[a] PCR 1,2,4,5,7 and 8 were all multiplex PCR reactions. PCR 3, 6 and 9 were all simplex PCR reactions.
[b] PCR primers were labelled with standard Beckman Coulter Dyes (D2, D3 and D4)

Amplification was carried out in 0.2µl thin-walled 96 well plates (Alpha, Eastleigh, UK) in a Perkin Elmer 9600 Thermocycler (Perkin Elmer Cetus, USA) for 3min at 95°C then 30 cycles of; 30 s at 95°C, 1 min at 60°C, 1 min at 72°C, and 1 cycle of 5 mins at 72°C.

2.5.2.2 Fragment Analysis

Fragment analysis software on the CEQ8000 capillary sequencer (Beckman Coulter, Fullerton, California) was used to determine the size of each VNTR-MIRU fragment. For VNTR-MIRU-12, two pools of 6 loci were generated. Pool A contained the PCR products from PCR reactions 1, 3 and 5 (table 2.5.2) and pool B contained the PCR products from PCR reactions 2, 4 and 6 (table 2.5.2). Each pool contained 2 loci labelled with one of the three Beckman Coulter

standard dyes (D2, D3 and D4). Each pool was run in separate capillaries on the sequencer. For VNTR-MIRU-15 three pools of 5 loci were generated. Pool A contained the 1,4 and 7, pool B contained PCR products from PCR reactions 2, 5 and 8 and pool C contained the PCR products from PCR reactions 3, 6 and 9. The Beckman Coulter dyes used to label each product are displayed in table 2.5.3. Again each pool was run on a separate capillary. The number of repeats per loci corresponded to the size of the amplified fragments displayed by the sequencer (see allele calling table – Appendix 3). PCR products labelled with the same dye could be distinguished because the primers were designed in such a way that the size of the repeats at each loci would never fall within 8 bp of each other. The final results were recorded as a 12 or 15 digit code.

2.6 Bionumerics Analysis

Analysis of all molecular fingerprints/profiles was performed using Bionumerics software (version 3.0 Applied Maths, Kortrijik, Belgium). Comparison of RFLP fingerprints (allowing a 3% tolerance) and spoligotypes was performed using the Dice similarity coefficient. All RFLP-IS*6110* fingerprints with an 80% similarity or greater were visually compared for potential clustering). VNTR-MIRU profiles were analysed using a categorical similarity coefficient. All dendrograms were displayed using UPGMA.

2.7 Hunter-Gaston Index

The Hunter-Gaston Index (HGI) determines the discriminatory power of a typing method. The HGI is based on the probability of two unrelated strains from a test population being placed into different typing groups. A result of 1 equals total

discrimination of the population, i.e. all isolates are unique (Hunter and Gaston 1988).

Figure 2.7.1 HGI formula

$$D = 1 - \frac{1}{N(N-1)} \sum^{s}{}_{j=1} nj\,(nj-1)$$

N = number of strains

S = number of different types

nj = number of strains belonging to type j

2.8 Macroarray analysis

Analysis of four lineage-defining silent single nucleotide polymorphisms (sSNP) within three genes (*oxyR, katG* and *rpoB)* was performed using a DNA macroarray. The development of the macroarray is described in chapter 6 and the final protocol is described below.

2.8.1 Amplification of target DNA

Biotin-labelled primers were designed to flank the four regions of interest and were based on the sequence of *M. tb* H37Rv and CDC1551 (Table 2.8.1). The reaction mix (total volume 21.2µl) contained 5µl sterile dH_2O, 10µl 2 x reaction buffer (3mM $MgCl_2$, 3.2mM dNTPs, 2x NH_4 buffer), 5µl of primer mix (2µM each primer), 0.2µl Taq polymerase (5U/µl), (Bioline, London, UK) and 1µl template DNA. Amplification was carried out in 0.2µl thin-walled 96 well plates (Alpha, Eastleigh, UK) in a Perkin Elmer 9600 Thermocycler (Perkin Elmer Cetus, USA) for 3 mins at 95°C then 30 cycles of 15 s at 95°C, 30 s at 60°C, 1 min at 72°C and one cycle of 5 min at 72°C.

Table 2.8.1 Primer sequences for the macroarray

Gene	Primers 5' to 3'*	Product size
*oxy*R	Forward - atcgccgccaagaggtgcta Reverse - tcacgcactgcacgacggt	311
katG	Forward - tgtcccgtcgtgggtcatat Reverse - ttgtccaagctggcgttgt	370
rpoB	Forward - tgcgtgtgtatgtggctcagaaa Reverse - cgccgtgggtgttcaaaataat	159
rpoB	Forward - gtaaggcgcagttcggtgg Reverse - tttgagcagcaccttgaacga	203

* All forward primers were 5' biotin labelled

2.8.3 Probe design

Two probes were designed per region of interest; one with the wild type sequence according to *M.tb* H37Rv and CDC1551, the other containing the sSNP. Each probe was between 19 and 25 nucleotides in length. A poly-T tail of 20bp was added to increase probe adhesion to the membrane (Table 2.8.2).

Table 2.8.2 Wild-type and mutant probe sequences

Gene & nt[a] position (sSNP)	Probe wt[b]	bp	Probe mutant	bp
oxyR-37 (G-A)	ccaccgcggcgaacgcgcgaag ccctttttttttttttttttt	35	gcggcgaacgcgcgaaaccc gttttttttttttttttttt	31
katG-87 (C-A)	accccgtcgagggcggcggttttt ttttttttttttt	29	acccagtcgagggcggcggat ttttttttttttttttt	30
rpoB-2646 (A-C)	cggccagcttgtcaccgtcggtttt ttttttttttttttt	31	cggccagcttgtccccgtcggt ttttttttttttttttt	31
rpoB-3243 (A-G)	ctcctgcagggtgtaggcagctttt ttttttttttttttt	31	tcctgcagggtgtaggcggcat ttttttttttttttttt	31

[a] nt – nucleotide position within gene. [b] Polymorphic site is highlighted in bold

2.8.4 Membrane preparation

Each probe stock (200μM) was diluted to a working concentration of 20μM in purified water containing 0.001% bromophenol blue. Each probe was dotted onto a nylon membrane (Magnagraph 0.22μM, Osmonics, Minnetonka, USA) using a hand-held arraying device (VP Scientific, San Diego, USA) and allowed to dry. Wild type and mutant probes were orientated adjacently on the membrane (Table 2.8.3). In addition an ink dot for orientation of the membrane and a colour control (2μl of *oxyR* primer) to check detection were added to the membrane. Each membrane was UV cross-linked at 12,000 $\mu J/cm^2$ for 1 min to fix the probes then washed twice in wash solution (0.5x SSC, 0.1% SDS) for 5 mins and allowed to air dry. The membranes were trimmed then placed in a 2ml plastic hybridisation tube and stored at room temperature.

Table 2.8.3 Membrane orientation

Ink dot	CDC[a]
Blank[b]	Blank
Blank	Blank
oxyR – 37	*oxyR* – 37mt[c]
katG - 87	*katG* – 87mt
Blank	Blank
Blank	Blank
rpoB - 2646	*rpoB* – 2646mt
rpoB - 3243	*rpoB* – 3243mt

[a] Colour development control

[b] Blanks were inserted in order to make the membrane longer for ease of handling.

[c] mt = mutant type

2.8.5 Hybridisation

Twelve microlitres of amplified *oxyR*-37 and *rpoB*-3243 from each sample were combined in a separate tube. K*atG*-87 and *rpoB*-2646 amplified products were first diluted 1 in 10, then 12μl was added to the mixture tube. The DNA mix was denatured at 100°C for 10 mins then added to 500μl of hybridisation solution (5xSSPE, 0.5% SDS) in a hybridisation tube containing a membrane. Each tube was placed on a rotisserie in a hybridisation oven at 60°C for one hour. Membranes were washed twice in 25ml of stringent wash buffer (0.3x SSPE, 0.5% SDS) for 1 min at room temperature then in 10 ml of preheated stringent wash buffer at 60°C for 10 mins.

2.8.6 Detection

All washes were carried out at room temperature. The membranes were washed twice in 25ml of rinse buffer (0.1M NaCl, 0.1M Tris-HCL, pH 7.5) for 1 min then incubated in 5ml rinse buffer with 0.5% blocking reagent (Roche, East Sussex, UK) (buffer should be stored at -20°C), before adding 20μl streptavidin-alkaline phosphatase (Roche, East Sussex, UK) and incubating for 30 mins. Each membrane was then washed twice in rinse buffer for 1 min and equilibrated for 1 min in buffer 3 (0.1M Tris, 0.1M NaCl, pH 9.5). The membranes were then incubated in colour development solution (15μl of p-Nitro Blue Tetrazolium Chloride (NBT) 75 mg/ml in 70% Dimethylformamide (DMF) and 15μl 5'-Bromo-4-Chloro-3-Indolyl phosphate (BCIP) 50 mg/ml in 100% DMF were added to 5ml of Buffer 3) for 1 hour. The streptavidin-alkaline phosphatase binds to the biotin in the forward primer. The colour development solution then

binds to this conjugate and produces a dark spot on the membrane indicating hybridisation between the PCR product and probe has taken place.

2.9 Single-stranded conformation polymorphism (SSCP)

SSCP using polyacrylamide gel electrophoresis (PAGE) was performed to identify potential sSNPs in *M. bovis*.

2.9.1 PCR analysis

Primers flanking the four target regions described in 2.9.1 were amplified under the same conditions. To denature the amplified DNA, 3µl of each product was added to the same volume of denaturing solution (400nM NaOH, 20mM EDTA pH 8.0) and left to denature for 5 mins. From this solution 2µl was then added to 2µl of loading buffer (Promega, Southampton, UK) and loaded onto the polyacrylamide gel. Two 100bp ladders and two controls were added per gel. A *M. bovis* strain previously characterised by sSNP analysis (Baker *et al.* 2004) was used as a wild-type control strain.

2.9.2 PAGE gels

A solution comprising of 4% glycerol (v/v) and 8% polyacrylamide (v/v) (40% acrylamide/bis solution, 29:1) made up to 50ml with 1 X TBE (0.1M Tris, 0.9M Boric Acid, 0.01M EDTA pH 8.4) was polymerised with 0.5% (w/v) fresh ammonium persulphate and 0.1% (v/v) tetramethylethylenediamine (TEMED) (All regents were from Bio-Rad, Hemel-Hampstead, UK). After polymerisation, the gel plates were loaded into a Protean II xi cell (Bio-Rad, Hemel-Hampstead, UK) containing 1 X TBE and run at 50V for approximately 16 hrs.

2.9.3 Silver-staining

Gels were fixed in a 40% (v/v) methanol for 30 mins followed by two washes in 10% (v/v) ethanol for 15 mins each. The gel was then stained with a Bio-Rad silver staining kit according to the manufacturer's instructions (Bio-Rad, Hemel-Hampstead, UK). Briefly, the gel was incubated in 200ml Bio-Rad oxidizing solution for 5 mins and rinsed twice in dH_2O for 5 mins. The gel was then incubated in 200ml Bio-Rad silver reagent for 20 mins and rinsed once with dH_2O. Following this, 200ml of Bio-Rad developing solution was added and incubated until bands appeared on the gel. The reaction was stopped with 5% (v/v) acetic acid.

2.10 DNA sequencing

DNA Sequencing was performed on a CEQ8000 capillary sequencer (Beckman Coulter, Fullerton, California). Amplified DNA from potential mutants previously identified in PAGE analysis was diluted 1 in 25 in sterile dH_2O. Sequencing reaction mix (total volume 5µl) contained; 2µl Quickstart master mix (Beckman Coulter, Fullerton, California), 2µl diluted DNA and 1µl of forward or reverse primer (primers were identical to those used for PAGE analysis). Each reaction underwent 30 cycles of 96°C for 20 sec, 55°C for 20 sec and 60°C for four mins. The labelled fragments were precipitated in a solution of 20µl sterile dH_2O, 2.5µl 3M NaOAc, 75µl ice cold ethanol and 2% (w/v) glycerol then spun at 1100 x g for 45 mins at 4°C. Pellets were washed twice at the same speed in 70% ethanol for 15 mins at 4°C. A paper towel was then placed on top of the open tubes. The tubes were inverted, placed into a centrifuge and pulse spun to

10 x g. The excess liquid was absorbed by the paper towel. The pellet was then dried in a desiccator for 15 mins, re-dissolved in 25µl sample loading solution (Beckman Coulter, Fullerton, California) and loaded onto the analyser. Sequences were analysed with CEQ analysis software.

2.11 Polyethylene glycerol (PEG) purification

PEG purification was performed on any isolates that would not sequence. Each amplified product was made up to 50µl with sterile dH_2O and incubated with 50µl of 20% PEG (8000 average molecular weight) /2.5M NaCl for 40 mins. The mixture was then spun at 9400 x g for 15 mins to pellet the DNA. The supernatant was discarded and the pellet washed twice with 150µl of 70% Ethanol and spun at 9400 x g for 5 mins. The supernatant was discarded and the pellet re-suspended in 20µl of sterile dH_2O.

CHAPTER 3.0 Molecular epidemiology of *Mycobacterium tuberculosis* infection in England, UK

3.1 Introduction

Rates of tuberculosis (TB) were declining in the UK, along with other industrialised countries, until the mid-1980s when the incidence of TB began to increase annually. This resurgence has been attributed to an increase in immigrants from high incidence countries, poverty, a failure in TB control procedures and improved national surveillance detecting the increased incidence (Hayward and Watson 1995).

In 2003 the incidence of TB in England and Wales was 12.8 per 100,000. In London the rate is much greater at 41.3 per 100,000. The national TB survey of England and Wales, completed every five years between 1978/9 and 1998, highlighted that the incidence of TB has increased at a higher rate in London and the West Midlands compared to low incidence areas such as Anglia and Oxford. Rates were especially high in people born outside the UK who had recently arrived in the UK (Kumar *et al.* 1997, Ormerod *et al.* 1998, Rose *et al.* 2001). In 1999 continuous enhanced surveillance was introduced whereby detailed clinical, epidemiological and microbiological information was collected on all TB cases in England and Wales and has confirmed the continuation of the increasing trend in TB incidence.

The advent of molecular typing techniques over the last decade has dramatically increased our understanding of TB transmission and dissemination. The

identification of related cases of TB has been beneficial in terms of directing clinical and public health action and for national surveillance.

The aims of this study were two fold. Firstly, to determine the molecular epidemiology of TB transmission in England in 1998 using a combination of molecular and epidemiological data, with particular regard to the level of recent transmission versus reactivation of disease. Secondly, to determine the role of molecular typing as a tool for the identification of clusters of TB that may be related for which clinical and public health action might be appropriate and for national surveillance. The results in this analysis represent a study of TB transmission in a city in a low incidence country. TB transmission in a city in a high incidence country, Harare, Zimbabwe is described in Chapter 4.0

3.2 Transmission of *M.tb* in England in 1998

3.2.1 Study population

This work was based on a collaborative study which assessed the molecular epidemiology of TB in England in 1998. All culture confirmed cases of TB in England between the 1st January and 31st December in 1998 were eligible for inclusion in the study. Epidemiological information was collected from three sources; laboratory records on each patient, the 1998 National TB survey of England and Wales and Mycobnet (Mycobacterial resistance network) a surveillance system for drug resistant isolates in the England and Wales. In addition, a questionnaire requesting further epidemiological information on

known risk factors was sent to the clinician in charge of each clustered case and also to a random selection of non-clustered cases in 1998.

3.2.2 Molecular analysis

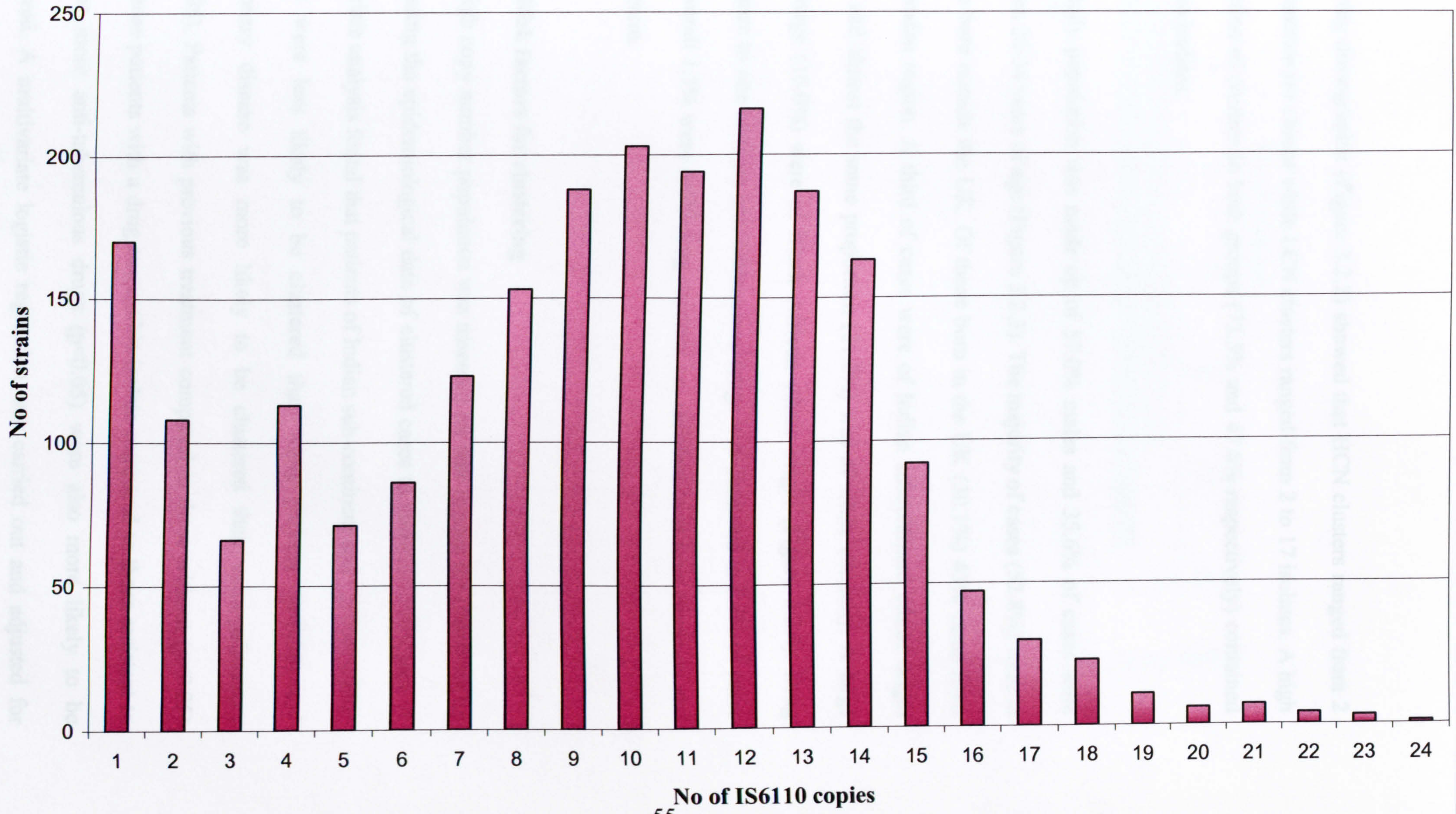
RFLP and spoligotyping (Material and Methods 2.3 and 2.4) were performed on all isolates from 1998. Clustered isolates, i.e. those with indistinguishable fingerprints, processed in the same hospital laboratory within 7 days were considered as possible laboratory cross-contamination and excluded from further analysis.

3.2.3 Results

In 1998, 3698 culture confirmed cases were identified of which 15 isolates were excluded due to possible laboratory cross contamination. Of the remaining isolates, 2265 (61.2%) fingerprints by both RFLP and spoligotyping were produced. RFLP fingerprints were divided into those containing 5 or more copies of the IS*6110* element i.e. high copy number (HCN) and those containing less than 5 copies i.e. low copy number (LCN) fingerprints. Fingerprints ranged from 1 to 24 bands with the majority (63.0%) of isolates containing between 7 and 14 bands (Figure 3.2.1).

The HCN group consisted of 1808 (78.9%) isolates of which 372 (20.6%) formed 152 clusters. Spoligotyping was applied as a second-line method to 457 isolates comprising the LCN group. A cluster in the LCN group was defined as those isolates containing indistinguishable RFLP and spoligotyping fingerprints. Based on this, 42 clusters containing 213 (46.6%) isolates were identified. The

Figure 3.2.1 Frequency of number of strains by number of IS6110 copies (1stJan -31st Dec 98)



clustering demographic (Figure 3.2.2) showed that HCN clusters ranged from 2 to 10 isolates per cluster while LCN clusters ranged from 2 to 17 isolates. A high proportion of clusters in both groups (73.3% and 47.6% respectively) contained just two isolates.

The study population was made up of 57.0% males and 35.6% of cases were between 20-34 years of age (Figure 3.2.3). The majority of cases (52.8%) were in people born outside the UK. Of those born in the UK (30.1%) 43% came from the London region. A third of cases were of Indian subcontinent ethnic origin (ISC) and almost the same proportion (31.0%) was of white ethnicity. A large percentage (16.0%) were of Black African ethnic origin (Figure 3.2.4). Drug resistance to one or more anti-tuberculous drugs was observed in 8.0% of cases and overall 1.5% were multi-drug resistant i.e. resistant to at least isoniazid and rifampicin.

3.2.4 Risk factors for clustering

The high copy number population was assessed for risk factors for clustering by comparing the epidemiological data of clustered cases to non-clustered cases. A univariate analysis found that patients of Indian sub-continent and Chinese ethnic origin were less likely to be clustered than white patients ($p<0.05$) and pulmonary disease was more likely to be clustered than extra-pulmonary ($p<0.05$). Patients with previous treatment compared to those without ($p<0.05$) and those patients with a drug susceptible isolate compared to those resistant to one or more anti-tuberculous drugs ($p<0.05$) were also more likely to be clustered. A multivariate logistic regression was carried out and adjusted for

Figure 3.2.2 Cluster demographic

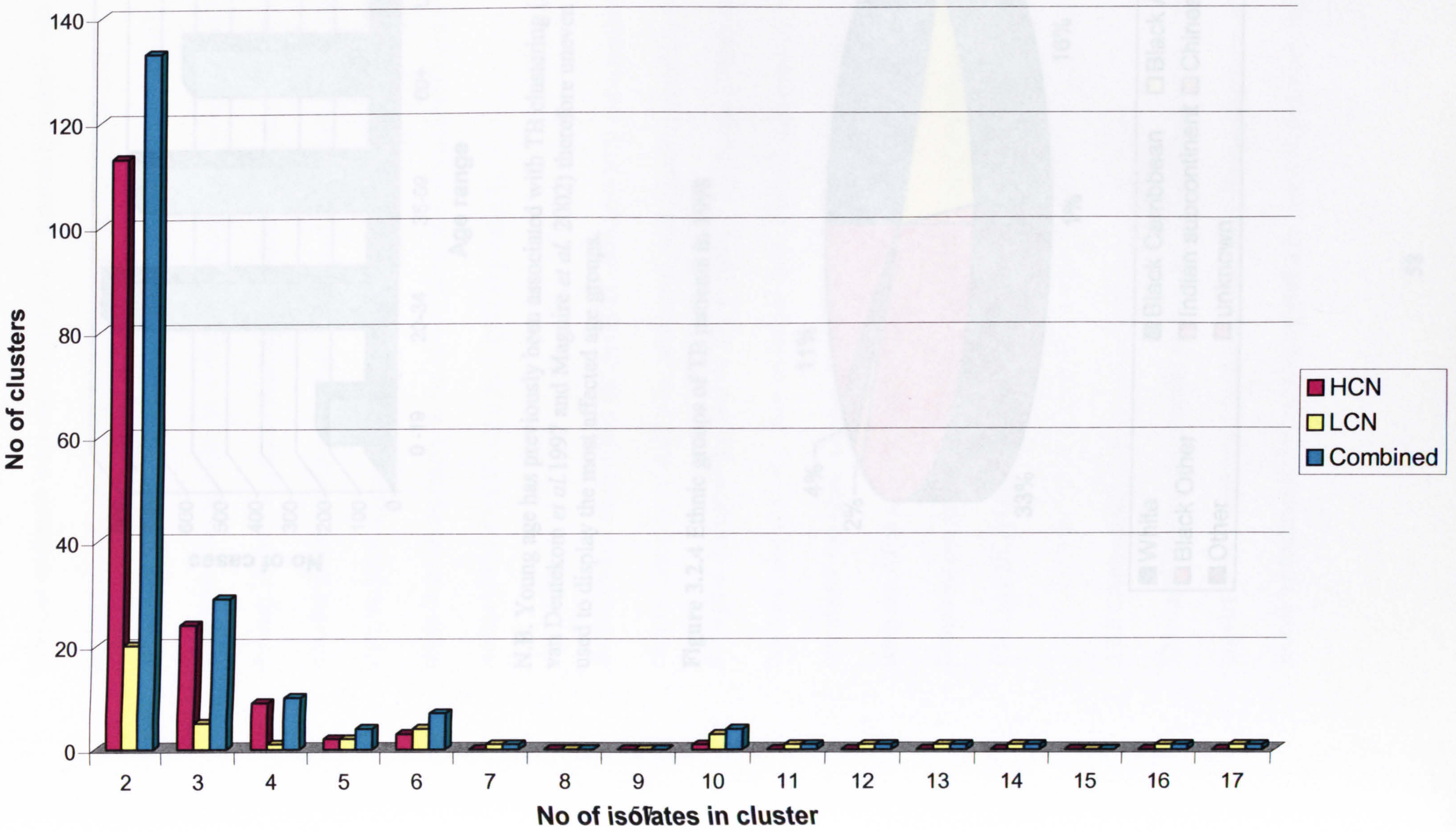
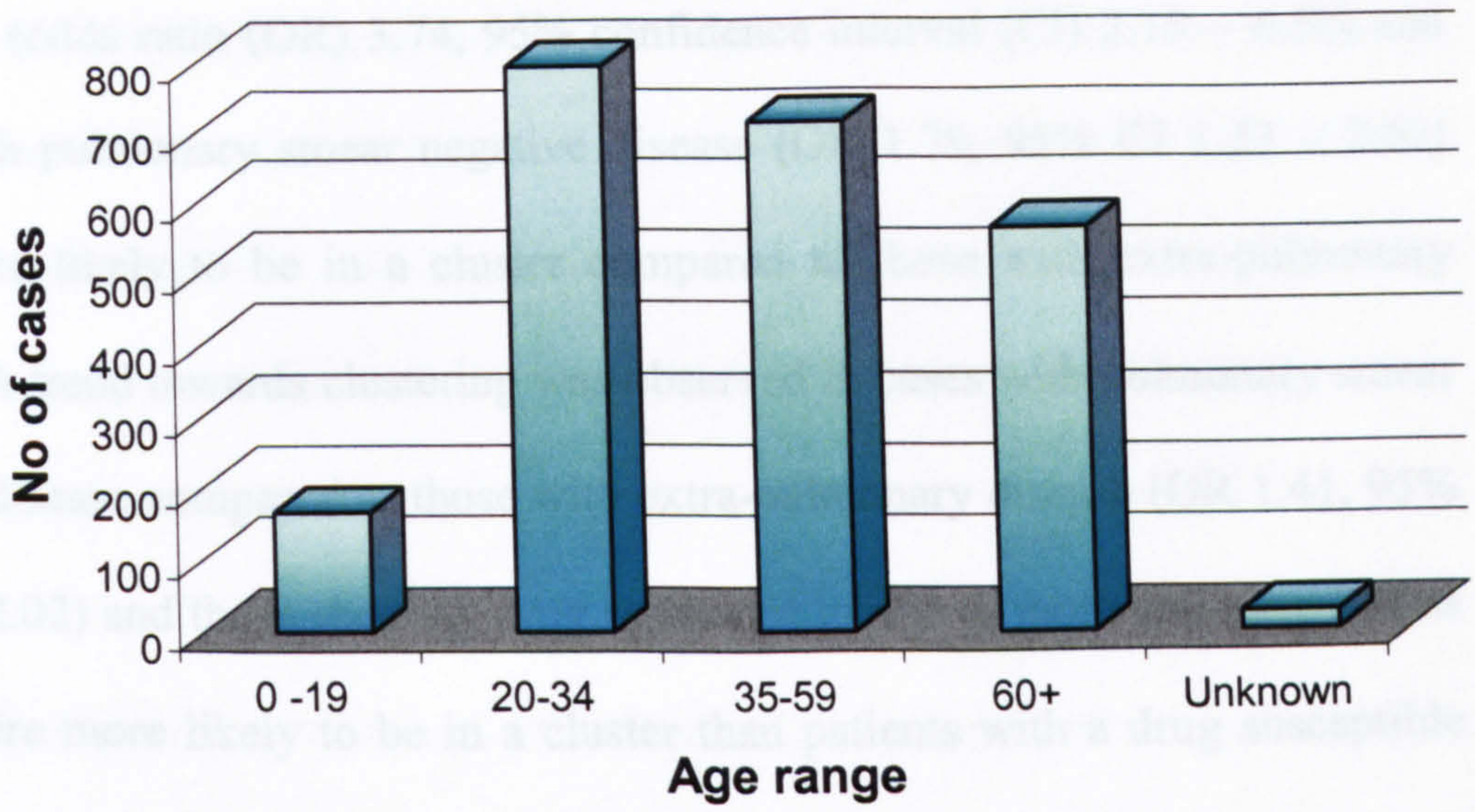
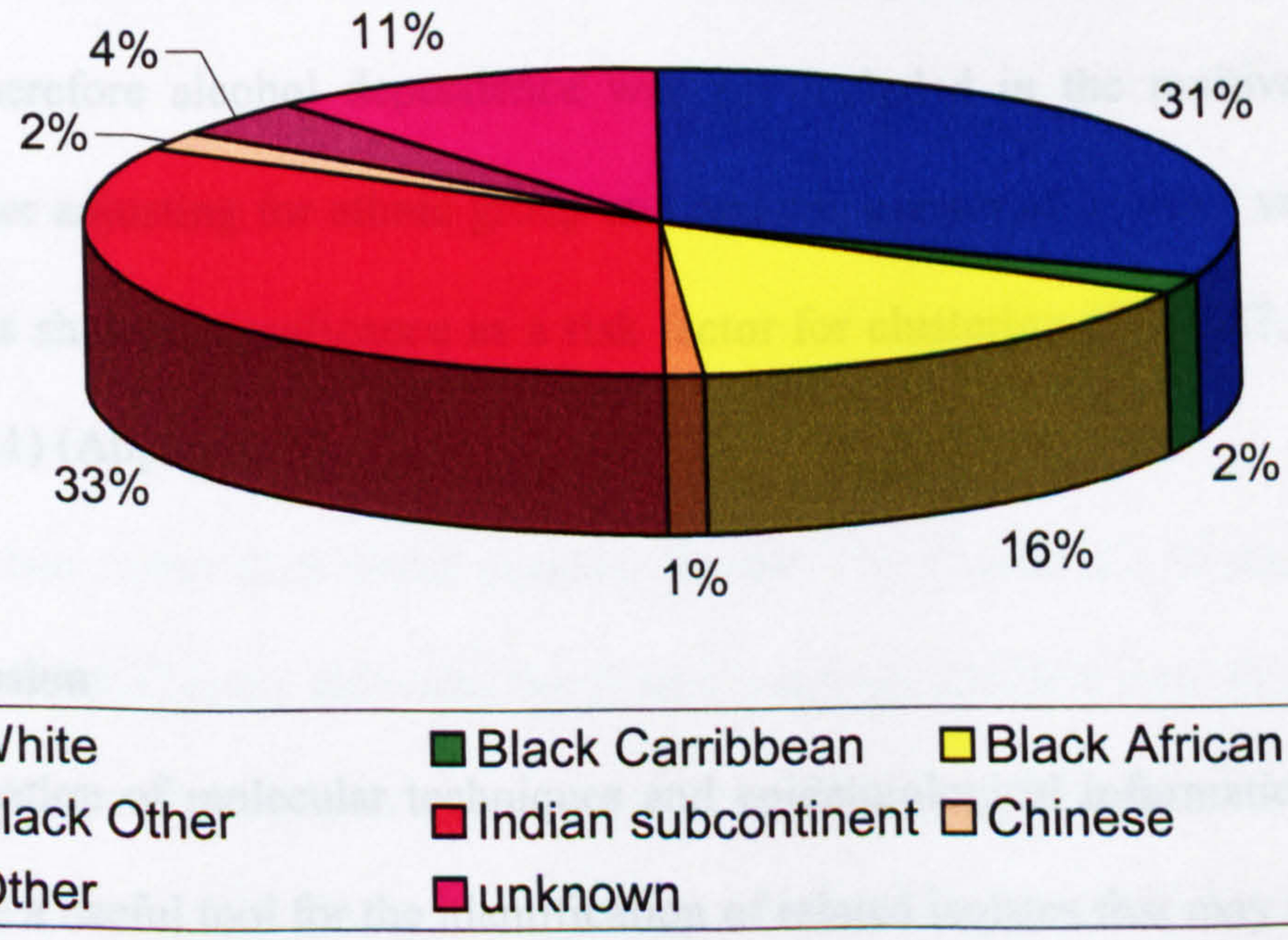


Figure 3.2.3 Age range of all TB patients in 1998



N.B. Young age has previously been associated with TB clustering (Alland *et al.* 1994, van Deutekom *et al.*1997 and Maguire *et al.* 2002) therefore uneven age intervals were used to display the most affected age groups.

Figure 3.2.4 Ethnic groups of TB patients in 1998



variables that were statistically significant in the univariate analysis. Age and gender were added *a priori*. Multivariate analysis showed that cases having had previous treatment were more likely to be clustered than those without previous treatment (odds ratio (OR) 3.74, 95% confidence interval (CI) 2.15 – 6.50) and those with pulmonary smear negative disease (OR 1.79, 95% CI 1.23 – 2.61) were more likely to be in a cluster compared to those with extra-pulmonary disease. A trend towards clustering was observed in cases with pulmonary smear positive disease compared to those with extra-pulmonary disease (OR 1.41, 95% CI 0.98-2.02) and those showing drug resistance to one or more anti-tuberculous drugs were more likely to be in a cluster than patients with a drug susceptible isolate (OR 1.43, 95% CI 0.93-2.19) (Appendix 4).

A case control study using all clustered cases (n = 372) and a random selection of control cases (n = 315) was performed. Additional information was collected through a questionnaire sent to the consultant in charge of each case. Bivariate analysis showed that homelessness and alcohol dependence were closely correlated therefore alcohol dependence was not included in the multivariate analysis. After adjusting for ethnic group and age and gender as *a priori* values, homelessness showed significance as a risk factor for clustering (OR 5.47, 95% CI 1.24 - 24.1) (Appendix 5).

3.2.5 Discussion

The combination of molecular techniques and epidemiological information has proven to be a useful tool for the identification of related isolates that may not be linked by traditional contact tracing methods alone. In this study the transmission

of TB in England in 1998 has been described using RFLP and spoligotyping alongside epidemiological data.

Clusters of identical fingerprints are thought to represent epidemiologically linked cases and therefore to reflect the level of recent transmission (Small *et al.* 1994, Alland *et al.*1994, van Deutekom *et al.* 1997, Maguire *et al.* 2002). The molecular analysis in this study suggests that the level of recent transmission in England appears to be relatively low.

The level of clustering observed (20.6% of HCN isolates) was comparable to a three year study performed in London between 1995-7 which found a clustering rate of 22.7% (Maguire *et al.* 2002) but lower then that observed in other developed countries; 28% in France (Torrea *et al.* 1996), 38% in Seville, Spain (Safi *et al.* 1997), 38% in New York city, US, (Alland *et al.* 1994), 40% in San Francisco, US (Small *et al.* 1994) and 46% in Amsterdam, The Netherlands (van Deutekom *et al.* 1997).

In this study cases with pulmonary TB were more likely to be clustered than those with extra-pulmonary disease. This is not surprising given that pulmonary TB represents an infectious case. However, clustering was linked to smear negative disease rather than smear positive disease. This maybe due to the short time window, linked cases including the source case may not have been detected. Smear negative, culture positive cases can account for a sizable amount of TB transmission (Behr *et al.* 1999) although they are less infectious they may take longer to diagnose leaving a longer period for transmission to occur.

Cases having had previous treatment were also associated with clustering. Reasons for this link are unclear. Perhaps this trend is due to incomplete treatment or reactivation of previously treated strains, which have survived in a dormant state but have reactivated and are then transmitted in the community. Previous TB treatment has been associated with cases having drug resistant TB (Conaty *et al.* 2004) but does not appear to have been previously identified as a risk factor for clustering in a predominantly drug sensitive population.

There was a trend towards clustering in cases having drug resistant TB. This is a cause for concern and needs to be monitored to ensure that drug resistant cases are controlled and their spread limited. This data is in contrast to previous studies showing in different populations that drug resistant strains are less likely to be clustered then susceptible stains as these strains are assumed to be less virulent (van Soolingen *et al.* 1999, Godfrey-Faussett et al 2000).

The case-control study showed that homelessness was linked to alcoholism and that homelessness was a risk factor for clustering. The health and living conditions associated with homelessness and alcohol dependence are well suited to the transmission of TB and have been previously been identified as risk factors for clustering (Kimerling *et al.* 1998).

It is conceivable that the level of ongoing transmission in this study may have been underestimated as the study duration was only one year, therefore contacts may not have been identified. Van Soolingen *et al.* (1999) showed that cluster size increases in accordance with increasing length of study time and begins to

plateau after approximately 3-4 years. Based on this curve the percentage clustering observed in the Netherlands at one year was approximately 33%. After five years the level of clustering had reached 46%, hence a 13% increase in clustering. Although the level of clustering in different populations is dependant on many factors including, demographic and clinical aspects, study design and typing methods utilised, the level of clustering in England based on this curve could reach approximately 34% in a five-year study. The interpretation of the DNA markers used in epidemiology studies depends on the stability of the regions analysed. Studies have demonstrated that RFLP-IS*6110* fingerprints are stable and have a half-life of between 3.2 and 8.7 years (de Boer *et al.* 1999, Warren *et al.*2002), therefore linked cases in this study are unlikely to be sub-divided by RFLP-IS*6110*.

A third of the culture confirmed cases in 1998 did not undergo molecular typing and in addition the majority of typed isolates were from the London and the south-east (63.9%) therefore the extent of clustering nationally may have been underestimated. Inevitably not identifying all epidemiologically linked cases would lead to an underestimate of the impact of risk factors for clustering.

Although clustering is a good indication of the level of recent transmission perhaps a more accurate method would be to use the n-1 approach whereby the index case is presumed to be a member of each cluster (Small *et al.*1994). Based on this the level of recent transmission in this study was 12.2%.

The majority of DNA fingerprints were unique. This could imply that TB disease in England is predominantly due to reactivation of a latent infection. A substantial proportion of cases will be due to reactivation of infection but importation of disease particularly from high incidence countries is likely to contribute a significant number as 52.8% of cases were born outside the UK and 12.8% of cases had resided in England for less than two years. As clinical TB commonly develops within two years of infection, these cases may reflect the level of importation of disease seen in England. A study of the origins of *M.tb* in London found isolates could be assigned to large superfamilies based on IS*6110* patterns, which were associated with country of birth. Transmission between immigrant and endogenous populations was found to be limited suggesting transmission occurred in the country of origin prior to entry into the UK or through close contact in certain communities (Dale *et al.* 2005).

A geographical analysis of clustering showed that clustered strains had disseminated throughout England demonstrating that TB transmission is not just a localised problem. The majority of clustered cases had not been identified previously through contact tracing. This supports the need for a national typing scheme as ongoing transmission cannot be monitored by local strain typing alone.

Molecular typing is an efficient method of identifying related cases however the interpretation of clusters must be treated with caution as false clusters with no epidemiological link may occur. Although it is possible that clusters with no apparent link may be the result of casual contact it is also possible that no link

exists. Sample size, study duration, incidence of TB and the stability of DNA regions analysed by a typing method can all effect clustering (Glynn *et al.* 1999, van Soolingen *et al.* 1999). To limit the amount of false clustering a typing technique providing a high level of discrimination is required. The discrimination of various typing techniques is discussed in the next section.

3.3 Comparative study of molecular typing techniques

3.3.1 Study population

The isolates described in section 3.2 and those with culture confirmed cases of TB in England between the 1st January 1999 and 31st March in 1999 were combined to assess the discriminatory power of different typing techniques.

3.3.2 Molecular Analysis

RFLP and spoligotyping (Material and Methods 2.3 and 2.4) were performed on all available isolates from this 15 month period (1st January 1998 to 31st March 1999). A geographical defined subset of this population (the London region comprised of 565 isolates) was selected and typed with VNTR-MIRU using 12 loci (VNTR-MIRU-12) in addition to RFLP and spoligotyping. A sub-set of this population (293 isolates) was chosen in a blinded but arbitrarily manner and typed with VNTR-MIRU using 15 loci (VNTR-MIRU-15) (Material and Methods 2.5.2). A comparison of strain typing methods was performed using the Hunter-Gaston index (Material and Methods 2.7).

3.3.3 Results

A total of 4767 culture confirmed cases were recorded in Mycobnet during the 15-month (1st Jan 1998 to 31st March 1999) study period. Overall 2688 isolates were successfully RFLP and spoligotyped.

The discriminatory power of each typing method, alone and in combination was determined using the Hunter-Gaston Index (HGI). A typing technique providing

good discrimination is one with a high HGI and small number of clustered strains in comparison to other techniques. The HGI determined that a combination of RFLP and spoligotyping gave the highest discrimination of the whole population (0.9997) than either technique alone (Table 3.3.1). RFLP gave a higher level of discrimination when used alone compared to spoligotyping (0.9982 and 0.9856 respectively). However, the discriminatory power of these two techniques differed considerably when analysing HCN and LCN isolates separately. Although RFLP was the most discriminating technique for HCN isolates, when used to analyse LCN isolates its discriminatory power is reduced (0.9998 to 0.9409). Conversely, spoligotyping performed more effectively than RFLP in a LCN population displaying a higher discriminatory power (0.9590 and 0.9409 respectively).

A comparison of the clustering observed in HCN isolates according to RFLP and spoligotyping showed that the majority of clusters (78.4%) were concordant i.e. clusters defined by RFLP had indistinguishable spoligotypes. A total of 33 clusters (containing 73 isolates) were not concordant i.e. a cluster by RFLP could be sub-divided by spoligotyping. Of these clusters the majority contained just two isolates (84.8%) and varied on average by 6 spacers, although over half (57.6%) varied by between 1 and 4 spacers. The high level of concordance observed suggests that these two techniques are not independent of each other, which has previously been observed in LCN isolates (Dale *et al.* 2003).

Table 3.3.1 A comparison of the discriminatory power of RFLP and Spoligotyping for the study population of 2688 isolates

Typing techniques	HCN (n = 2152)					LCN (n = 536)					HCN + LCN
	n of clusters	Total n of types	n of clustered isolates (%)	n of unique types	HGI	n of clusters	Total n of types	n of clustered isolates (%)	n of unique types	HGI	HGI
RFLP	153	1873	432 (20.1)	1720	0.9998	44	141	439 (81.9)	97	0.9409	0.9982
Spoligotyping	177	421	1908 (88.7)	244	0.9743	43	173	406 (75.7)	130	0.9590	0.9856
RFLP + Spoligotyping	119	2010	357 (16.6)	1922	0.9999	53	325	264 (49.2)	272	0.9919	0.9997

A geographically defined population subset of TB isolates from patients residing in London region (n = 565) was tested using 12-loci VNTR-MIRU (VNTR-MIRU-12). The results are shown in Table 3.3.2. By comparing the three techniques when used alone, VNTR-MIRU-12 had a comparable level of discrimination compared to RFLP-IS*6110* (0.9958 to 0.9917 respectively) based on the maximum number of sub-divisions produced. Interestingly, the discriminatory power of VNTR-MIRU-12 was higher when analysing LCN isolates than that of RFLP and spoligotyping (0.9894 compared to 0.9101 and 0.9691 respectively). In the HCN population RFLP and VNTR-MIRU-12 were both highly discriminatory compared to spoligotyping. RFLP and VNTR-MIRU-12 had similar discriminatory values however the number of clustered isolates was far greater when the population was analysed with VNTR-MIRU-12 compared to RFLP (196 and 42 isolates respectively, i.e. less discriminatory).

In practice, RFLP is often used alone for isolates with a HCN in population studies and is then combined with a secondary typing technique in order to differentiate LCN isolates. In the total population a combination of RFLP and VNTR-MIRU-12 gave a similar level of discrimination as RFLP and spoligotyping according to the HGI (0.9997 and 0.9989 respectively). However the number of clustered isolates varied considerably in the HCN and LCN populations. The number of clustered isolates in the HCN population produced by RFLP and VNTR-MIRU-12 (n = 24) was lower than RFLP and spoligotyping (n =31). In the LCN number population the difference in the level of clustering was more pronounced between RFLP and VNTR-MIRU-12 (n = 49) and RFLP and spoligotyping (n = 70). Therefore, although the HGI of both appears to be

Table 3.3.2 A comparison of the discriminatory power of RFLP, Spoligotyping and VNTR-MIRU-12 for a population of 565 isolates

Typing Techniques	HCN (n=394)					LCN (n=171)					HCN +LCN
	n of clusters	Total n of types	n of clustered isolates (%)	n of unique types	HGI	no of clusters	Total n of types	n of clustered isolates (%)	n of unique types	HGI	HGI
RFLP	18	370	42 (10.7)	352	0.9996	12	53	130 (76.0)	41	0.9101	0.9917
Spoligotyping	43	180	257 (65.2)	137	0.9738	17	81	107 (62.6)	64	0.9691	0.9824
VNTR-MIRU-12	53	251	196 (49.7)	198	0.9945	24	115	80 (46.8)	91	0.9894	0.9958
RFLP + Spoligotyping	13	376	31 (7.9)	363	0.9997	17	118	70 (40.9)	101	0.9889	0.9989
RFLP + VNTR-MIRU-12	11	158	24 (6.1)	147	0.9998	19	364	49 (28.7)	345	0.9969	0.9997
VNTR-MIRU-12+ Spoligotyping	29	333	90 (22.8)	304	0.9982	20	138	53 (31.0)	118	0.9960	0.9987
RFLP + VNTR-MIRU-12+ Spoligotyping	9	384	19 (4.8)	375	0.9999	11	156	26 (15.2)	145	0.9986	0.9999

high the difference in the level of clustering seen appears to suggest that a combination of RFLP and VNTR-MIRU-12 gives a higher level of discrimination.

The greatest discrimination was achieved by using all three typing methods together. The discriminatory power of RFLP and VNTR-MIRU-12 was similar to that of all three techniques combined (0.9997 and 0.9999 respectively). Nevertheless, the level of clustering observed suggests three techniques were more discriminatory: clustering was reduced from 24 to 19 in the HCN population and from 49 to 26 in the LCN population by applying RFLP and VNTR-MIRU-12 and all three techniques respectively.

A sub-set (293 isolates) from the 565 isolates were chosen in a blinded but arbitrarily manner. These isolates were further analysed with RFLP-IS*6110*, VNTR-MIRU-12 and in addition VNTR-MIRU-15. Overall, all three techniques were highly discriminating (Table 3.3.3). VNTR-MIRU-12 and -15 typing performed better in the LCN population than RFLP typing (0.9303, 0.9888, and 0.9940 respectively). In general, RFLP is poor at differentiating between LCN isolates. This was demonstrated by the number of clustered isolates observed in each group ranging from; 55 with RFLP typing to 25 with VNTR-MIRU-15 analysis. A substantial proportion of isolates in this population were LCN (30.0%) therefore this will have acted to reduce the HGI of RFLP further. Conversely, in the HCN population the number of clustered isolates produced by RFLP was much lower than VNTR-MIRU-12 and VNTR-MIRU-15 (n = 16, 95,

Table 3.3.3 A comparison of the discriminatory power of RFLP, VNTR-MIRU-12 and -15 for a population of 293 isolates

Typing techniques	HCN (n=205)					LCN (n=88)					HCN + LCN
	No of clusters	Total no of types	No of clustered isolates (%)	No of unique types	HGI	No of clusters	Total no of types	No of clustered isolates (%)	No of unique types	HGI	HGI
RFLP	8	197	16	189	0.9996	8	41	55	33	0.9303	0.9936
VNTR-MIRU-12	28	138	95	110	0.9919	9	67	30	58	0.9888	0.9943
VNTR-MIRU-15	24	146	83	122	0.9936	10	73	25	63	0.9940	0.9960
RFLP + VNTR-MIRU-12	6	199	12	193	0.9997	5	83	11	77	0.9982	0.9997
RFLP + VNTR-MIRU-15	6	199	12	193	0.9997	5	83	11	77	0.9982	0.9997

83 respectively) demonstrating that the discriminatory power of RFLP surpasses VNTR-based typing in HCN populations.

3.3.4 Discussion

Determining the discriminatory power of a strain typing technique permits the identification of the best strain-defining technique or techniques for understanding how TB has been transmitted within a population. The rapid identification of related TB cases is paramount for efficient clinical and public health action but is also crucial for national surveillance so that resources can be targeted to problem areas.

In this study we used a population-based study to compare the discriminatory power of different typing techniques. A combination of RFLP and spoligotyping proved to be more discriminating compared to each technique alone. When both techniques were used alone there were variations in their discriminatory power depending on the composition of the population e.g. in the LCN population the discriminatory power of RFLP was reduced whereas spoligotyping was very effective at sub-dividing LCN clusters. This is consistent with previous studies which have shown a first and second line typing approach for LCN isolates provides a better level of discrimination whereby RFLP is used to analyse all isolates followed by spoligotyping of LCN cases to sub-divide RFLP LCN clusters (Goyal *et al.* 1997, Kamerbeek *et al.* 1997). However, caution should be taken in interpreting RFLP and spoligotyping clusters within LCN isolates to estimate the extent of recent transmission in epidemiological investigation as it

has been shown that there is congruence between these techniques which suggests distinct families of strains exist (Dale *et al.* 2003).

An analysis of the HCN population by both RFLP and spoligotyping demonstrated that the clusters produced by these techniques were not always concordant. Although RFLP is regarded as the gold standard for typing *M.tb*, in this study spoligotyping subdivided HCN clusters which could have implications for molecular epidemiological studies.

RFLP IS*6110* elements are believed to cause changes in the genome by transposition. Within the direct repeat (DR) region, on which spoligotyping is based, there is a copy of the IS*6110* element which can alter the composition of this region by removing direct variant repeats (DVR) during transposition (Fang *et al.* 1998). Evolution of this region can also be IS*6110*-transposition independent by the deletion of single or contiguous stretches of DVRs through homologous recombination of neighbouring or distant DRs (van Embden *et al.* 2000). Less than a quarter of HCN RFLP clusters in this study did not have indistinguishable spoligotypes. The molecular clock of RFLP is believed to be faster then that of spoligotyping (Niemann *et al.* 1999, Warren *et al.* 2002) therefore any difference in the latter may be seen as a more substantial biological change. However, taking this into account it is unlikely that those HCN clusters sub-divided by spoligotyping are false clusters and that all clusters produced by spoligotyping in the HCN population are the result of true transmission events. RFLP has proven to be a robust technique for strain typing as numerous studies have proven epidemiological links within RFLP HCN clusters (van Soolingen *et*

*al.*1999, Maguire *et al.* 2002, Ruddy *et al.* 2004). Furthermore a previous study demonstrated that spoligotyping can indeed split up epidemiologically linked HCN RFLP clusters (Kwara *et al.* 2003).

In our study the discriminatory power of spoligotyping was lower than that of RFLP in HCN isolates and in addition spoligotyping clusters have combined HCN and LCN isolates which are unlikely to be epidemiologically linked. The reasons behind the discrepancies between these techniques are unclear and may be a result of the mechanisms controlling their position and/or presence or absence.

Several studies have shown that VNTR-MIRU has a discriminatory power similar to RFLP (Supply *et al.* 2001, Cowan *et al.* 2002, Hawkey *et al.* 2003) however Nguyen *et al.* (2004) found that the discrimination provided by VNTR-MIRU was still lower than RFLP. In this study VNTR-MIRU typing provides a higher level of discrimination than spoligotyping for both HCN and LCN isolates. Although the HGI value of VNTR-MIRU was comparable to RFLP the former actually clusters more isolates compared to the gold standard. The additional clusters produced by VNTR- MIRU in the HCN population could be described as false clusters. A probable reason for the similar HGI values of these two techniques could be due to the size of the clusters. A technique producing a large number of very small clusters can inflate the HGI and was observed with VNTR-based typing which gave the impression that its discriminatory power was comparable to RFLP. The value of VNTR-MIRU typing may also have been overestimated in the final population (n = 293), where 30% of cases were LCN,

because the value of RFLP will decrease with an increasing number of LCN isolates. VNTR based typing does perform better than RFLP in LCN populations, seen in this study and with other work performed in our department (Gopaul *et al.* unpublished), but in HCN populations RFLP remains the most discriminatory technique.

A previous study demonstrated that the discrimination of VNTR-MIRU-12 could be improved by the addition of three loci from the Frothingham and Meeker O'Connell study; ETR-A, ETR-B, and ETR-C and that this 15 loci panel gives rise to discrimination that is similar to RFLP-IS*6110* analysis alone (Hawkey *et al.,* 2003). This suggests that a prospective typing service could be run as amplification techniques are more rapid and require much less DNA compared to RFLP which is a slow cumbersome technique requiring viable cultures. However the numbers employed in this study were small and analysis was based on selected investigations with epidemiological information. In our study there was an improvement in the discrimination of VNTR-MIRU with the addition of ETR-A to C. However there is still a large difference between the level of clustering produced by RFLP and VNTR-MIRU analysis. This is in accordance with further work performed in our department (Gopaul *et al.* unpublished).

The role of molecular typing techniques for public health benefit and for national surveillance is undoubtedly important but the methods for achieving the best results are ambiguous. No single technique provides the level of discrimination required for confirming or disproving epidemiological investigations. Whilst VNTR-MIRU is a promising technique, RFLP would still be required as a

primary screen for isolates from the UK due to the higher level of discrimination produced with HCN strains which make up the majority of cases. In terms of LCN isolates the use of RFLP-IS*6110* and VNTR-MIRU provides a better level of discrimination over that observed with the frequently used strategy of RFLP and spoligotyping.

Note:

This study was part of a collaboration between the HPA-Mycobacterium Reference Unit and the communicable disease surveillance centre (CDSC) and was funded by the Department of Health. Andrea Gibson performed all of the spoligotyping analysis and some of the RFLP and VNTR-MIRU. The RFLP analysis was mainly performed by Dr. Zack Fang (1998 isolates) and Dr. Krishna Gopaul (1999 isolates). The remaining VNTR-MIRU analysis was performed by Dr. Krishna Gopaul. Epidemiological data was analysed mainly by CDSC and in particular by Dr. Jane Love. Ethical approval for this study was granted by the Multi-centre Research Ethics Committee (ref. MREC/99/2/84)

CHAPTER 4.0 Molecular epidemiology of *Mycobacterium tuberculosis* in Harare, Zimbabwe.

4.1 Introduction

The incidence of Tuberculosis (TB) in Zimbabwe has risen dramatically over the last two decades from 60 per 100,000 in 1981 to 384 per 100,000 in 2001 according to the WHO (www.stoptb.org/events/world_tb_day/2002/News_Articles/_AFP_020322.htm). This resurgence, as in other parts of Africa has largely, been fuelled by the HIV epidemic in sub-Saharan Africa (Mwinga and Bernard Fourie 2004).

The gold-standard for typing *Mycobacterium tuberculosis* (*M.tb*), restriction fragment length polymorphism (RFLP), is a costly, time consuming technique which requires viable cultures and tends to be impractical for resource poor settings as suitable culture facilities are not always available. PCR techniques, such as spoligotyping and variable number tandem repeat (VNTR) (Kamerbeek *et al.* 1997, Frothingham and Meeker-O'Connell 1998), are more suited to this kind of environment as PCR can be performed directly on clinical samples and therefore viable cultures are not necessarily required. Spoligotyping and VNTR do not provide the same level of discrimination as RFLP in strains with more than four copies of IS*6110* as they can link together isolates with different RFLP patterns, but when applied together their discrimination of *M.tb* populations is improved (Goyal *et al.* 1997, Kremer *et al.* 1999, Filliol *et al.* 2000).

The aims of this study were to investigate the level of recent transmission versus reactivation of TB in Harare, Zimbabwe using two molecular typing techniques, spoligotyping and VNTR-exact tandem repeats (VNTR-ETR) and secondly to identify any spoligotyping families present in this high incidence population and compare them to those seen in a low incidence population in London, UK.

4.2 Study design

4.2.1 Patients enrolled on the study

A total of 516 consecutive smear positive patients presenting at the Beatrice Road Hospital in Harare between May and October 1997 were enrolled in the study. The Beatrice Road TB Hospital is the main referral centre for TB and other infectious diseases. Each patient was given a clinical examination, chest X-ray and a HIV test. A total of three sputum samples were collected per patient. Sputa were stained using a carbol fusion based stain and smear positivity (i.e. most infectious) and negativity were recorded. Demographic data and medical history of each patient was collected using a standardised questionnaire administered by a trained nurse.

4.2.2 Microbiological methods

Sputa were all cultured on Lowenstein-Jenson media. Culturing of sputum samples and drug susceptibility testing was performed at the TB reference laboratory, Bulawayo, Zimbabwe. HIV status was also determined in Zimbabwe using an ELISA (enzyme-linked immuno sorbent assay) screening test (Dupont, Wilmington, Del.). CD4 cell counts were measured by flow cytometry (FACScan Becton Dickinson, Paramus, N.J.)

4.2.3 Molecular methods

Cultures were transferred from Zimbabwe to the HPA Mycobacterium Reference Unit in London. A crude DNA extract (Materials and Methods 2.2.1) from each available culture was amplified and analysed by spoligotyping and VNTR-ETR (Materials and Methods 2.4 and 2.5.1). The discriminatory power of these two techniques was calculated using the Hunter Gaston Index (Material and Methods 2.7)

4.2.4 Laboratory Contamination

Samples processed in the microbiology laboratory on the same day as a positive smear from another patient with the same typing pattern were considered as laboratory cross contamination and excluded from the study.

4.3 Results

4.3.1 Molecular analysis

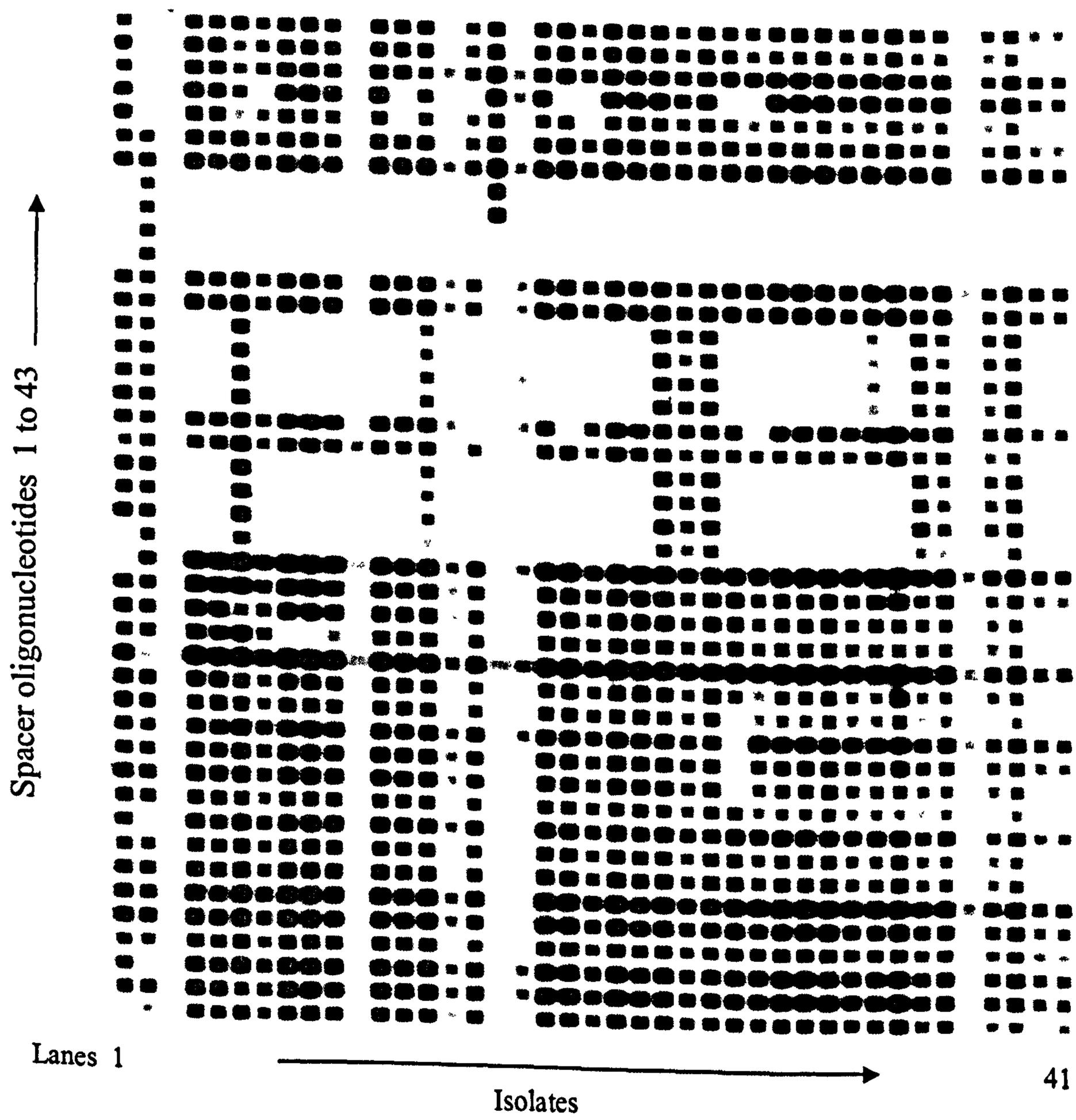
Of the 516 patients enrolled on the study, 502 had complete epidemiological and microbiological data and were included in the analysis. DNA was extracted from 224 cultures from 502 patients (44.6%). The remaining cultures were either non-viable, heavily contaminated with bacteria or yeast or the glass vials containing the cultures were broken during transportation from Harare to London rendering them unusable.

A total of 214/224 (95.5%) isolates were spoligotyped from which 54 distinct spoligotypes were produced. The remaining ten isolates failed to amplify. These spoligotypes produced 20 clusters (180 isolates) and 34 unique types. A cluster was defined as two or more isolates with indistinguishable fingerprint patterns. Cluster size ranged form 2 to 68, with the majority of clusters (85.0%) containing 8 or less isolates. Three clusters were disproportionately larger containing 23, 25 and 68 and accounted for 54.2% of the isolates. Examples of the two largest clusters are seen in figure 4.1.

VNTR-ETR typing was successfully performed on 198/224 (88.4%) isolates and again the remaining isolates failed to amplify. A total of 33 VNTR-ETR profiles were produced containing 12 clusters (177 isolates) and 21 unique VNTR-ETR types. Of the clustered isolates, 140 isolates (70.7%) formed just two clusters containing 127 and 13 isolates while the remaining clusters contained less than 7 isolates.

The discriminatory value of spoligotyping (0.8708) was higher then that of VNTR-ETR (0.5827). By combining the two typing techniques the discriminatory power was increased to 0.9221. Spoligotyping as a first line typing technique followed by VNTR-ETR sub-divided the 20 spoligotyping clusters into 24 clusters containing between 2 and 48 isolates (Table 4.1). The largest spoligotype cluster, containing 68 isolates, was divided into 8 different VNTR-ETR types. Overall 25 isolates had spoligotyping data but were unable to

Figure 4.1 Examples of the spoligotypes observed in Harare, Zimbabwe



Important features;
Lanes 1 and 2 – *M.tb* H37Rv and *M. bovis* BCG positive controls
Lanes 4 and 5 - examples of the spoligotyping pattern of the largest cluster (n = 68)
Lanes 24 to 26 – examples of the spoligotyping pattern of the second largest cluster (n = 25)

Table 4.1 Spoligotyping and VNTR used as a first and second line method

First-line typing method		Second-line typing method		
Spoligotyping		VNTR		
Cluster size	Fingerprint	No of sub-clusters	No of isolates per cluster	No of unique isolates[a]
4	000000000003771	0	0	1 (3)
2	757777777413731	0	0	1 (1)
68	777777606060771	6	2, 2, 2, 2, 48, 3	2 (7)
2	770077606060731	0	0	1 (1)
23	777777607760771	2	2, 16	2 (3)
3	477777607760771	1	2	1
3	477777777760771	1	2	0 (1)
5	777777677760771	2	2, 3	0
8	777417606060731	1	8	0
4	777737777760771	1	2	1 (1)
2	677777607760771	0	0	2
2	777777774020731	1	2	0
3	777775606060771	1	3	0
2	703377400001771	0	0	1 (1)
7	777777606060731	1	7	0
3	777777606060571	1	3	0
25	777777777760771	3	4, 2, 8	7 (3)
4	777777606060631	1	4	0
3	777777606060671	1	2	1
7	777777606060731	1	7	0
VNTR		Spoligotyping		
Cluster size	Fingerprint	No of sub-clusters	No of isolates per cluster	No of unique isolates
5	32333	2	2, 2	1
4	21232	2	2, 2	0
4	21233	2	2, 2	0
3	21431	3	2	1
5	21432	1	2	3
127	21433	13	48, 16, 2, 3, 8, 2, 3, 7, 3, 4, 4, 2, 7	10 (8)
2	42435	0	0	1 (1)
13	32433	1	8	5
2	22432	0	0	2
6	22433	2	2, 3	1
3	23433	0	0	2 (1)
3	32332	0	0	3

[a] Figures in parentheses could not be typed by the second-line typing method.

be VNTR-ETR typed therefore the value of VNTR-ETR may have been underestimated.

Alternatively, VNTR-ETR as a first line technique followed by spoligotyping sub-divided the original 12 clusters into 24 clusters containing between 2 and 48 isolates. The largest VNTR-ETR cluster, which contained 127 isolates, was sub-divided into 23 different spoligotypes.

None of the clustered isolates were processed in the laboratory on the same day and therefore the risk of cross-contamination was reduced.

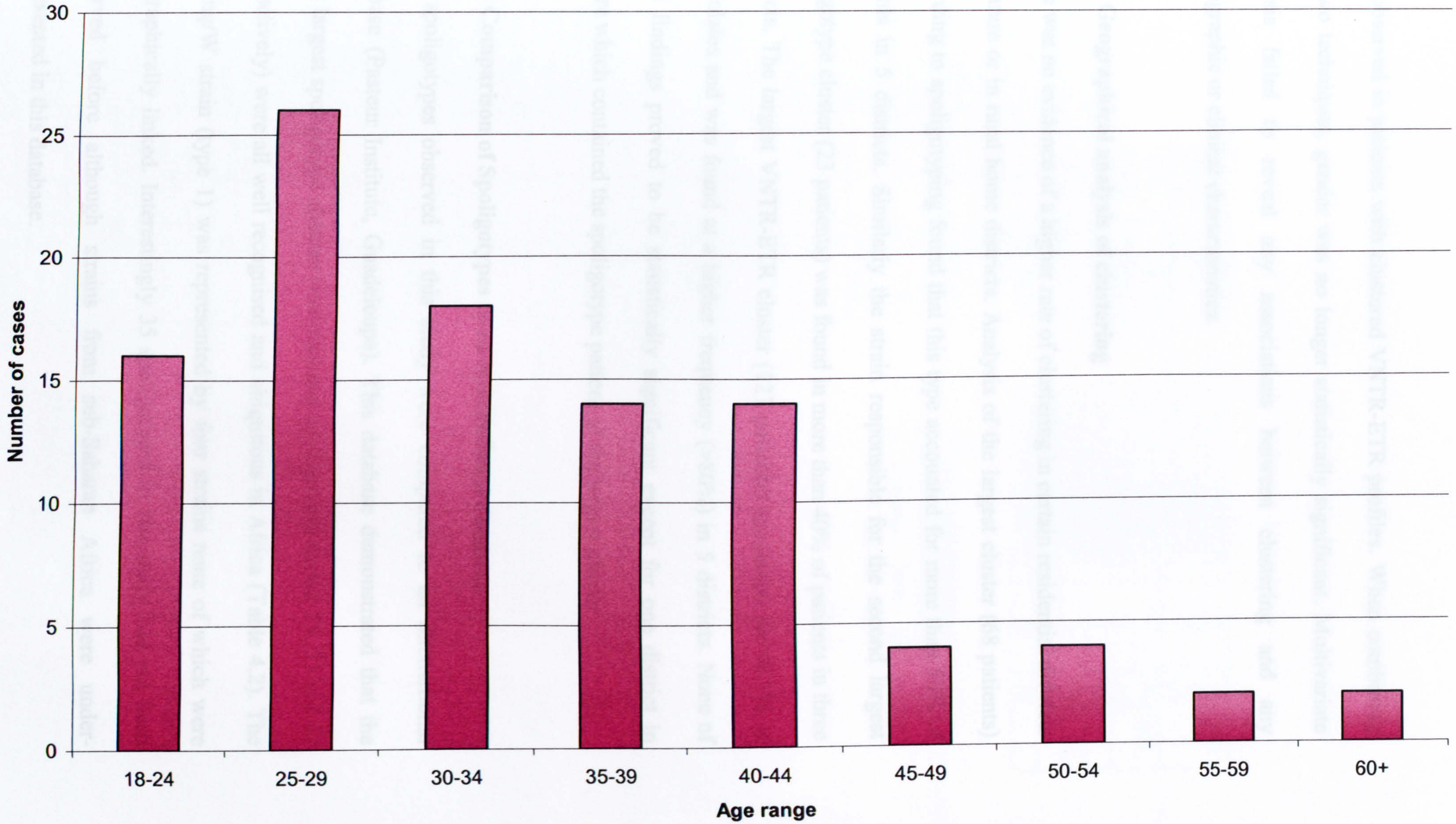
4.3.2 Epidemiological information

Of the 502 patients, 326 (64.9%) were male and the majority of patients were aged between 25-29 years with 43.6% of cases aged 25 to 34 years (Figure 4.2). Forty-four (8.9%) had a past history of TB and 204 (40.6%) gave a history of recent household contact with TB. A high proportion of patients (73.9%) were HIV positive and 17/458 (3.7%) of cases were resistant to one or more antimicrobial drugs.

4.3.3 Risk factors for clustering

Risk factors for clustering of patients' isolates were identified by comparing the epidemiological data of patients in clustered and unclustered groups using a chi-squared test. Univariate analysis of patients clustered by spoligotyping found patients with isolates in a cluster were less likely to be male ($p = 0.02$), this was

Figure 4.2 Number of cases in each age group



also observed in patients with clustered VNTR-ETR profiles. When combining the two techniques, gender was no longer statistically significant. Multivariate analysis failed to reveal any associations between clustering and any demographic or clinical characteristics.

4.3.4 Geographical analysis of clustering

There was no evidence of a higher rate of clustering in certain residential districts in Harare or in rural home districts. Analysis of the largest cluster (68 patients) according to spoligotyping found that this type accounted for more than 50% of patients in 5 districts. Similarly the strain responsible for the second largest spoligotype cluster (23 patients) was found in more than 40% of patients in three districts. The largest VNTR-ETR cluster (127 patients) accounted for 64.1% of all isolates and was found at a higher frequency (>80%) in 5 districts. None of these findings proved to be statistically significant except for one district in Harare which contained the spoligotype pattern of the largest cluster.

4.3.5 Comparison of Spoligotypes to an international database

The spoligotypes observed in this study were compared to an international database (Pasteur Institute, Guadeloupe). This database demonstrated that the three largest spoligotype clusters (international spoligotypes codes; 59, 53 and 42 respectively) were all well recognised and ubiquitous in Africa (Table 4.2). The Beijing/W strain (type 1) was represented by four strains none of which were geographically linked. Interestingly 35 spoligotypes in this study had not been observed before although strains from sub-Saharan Africa were under-represented in this database.

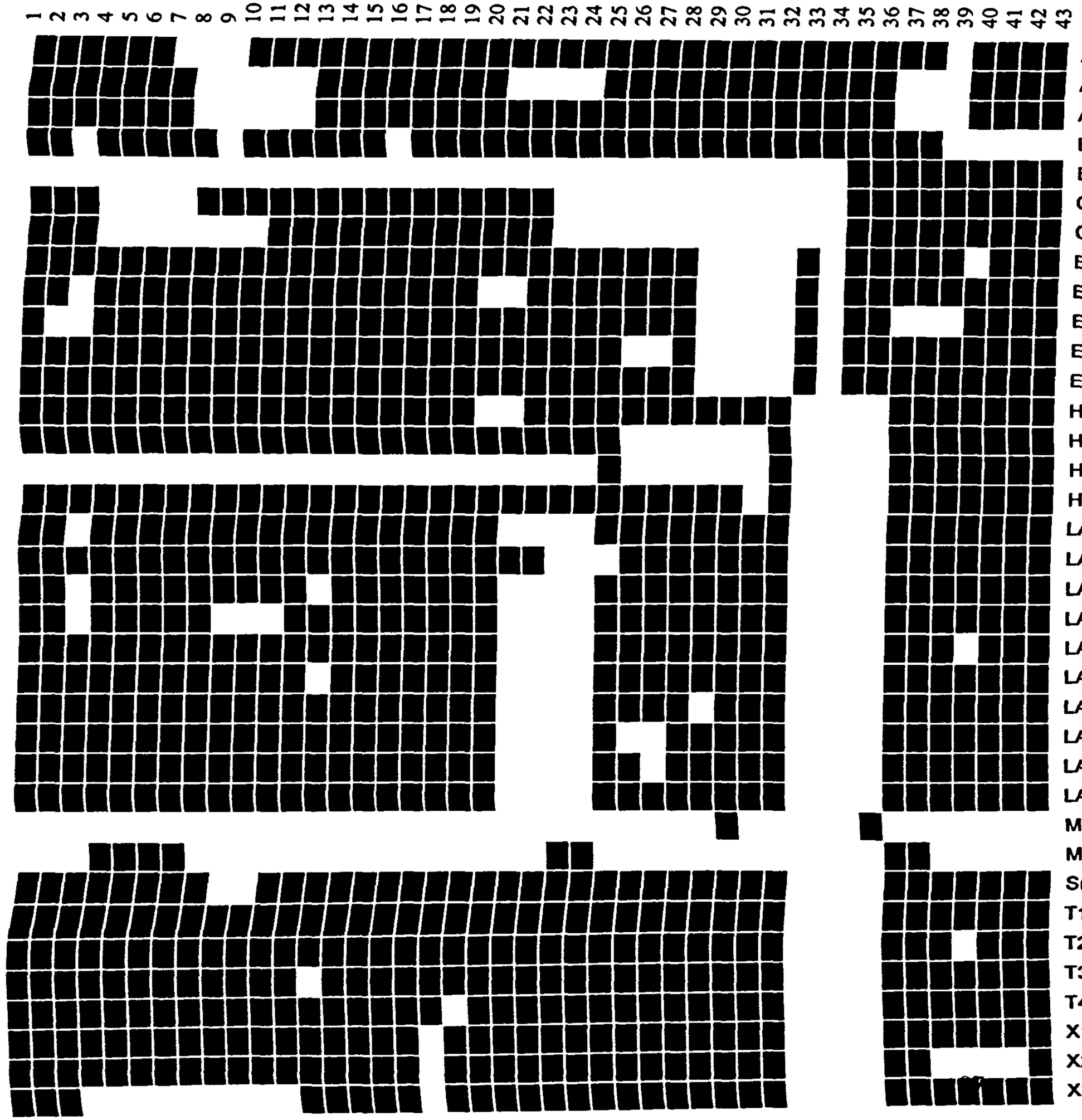
Table 4.2 Frequency of spoligotypes in Harare, Zimbabwe

Octal code[a]	No of isolates observe-d in Harare	Internati-onal spoligoty-pe database code[b]
000000000003771	4	1
077777606060771	1	816
100775747413771	1	Not seen
477777607760771	3	753
477777777760771	3	804
577777603760771	1	Not seen
601777606060731	1	Not seen
637775777763770	1	Not seen
677777607760771	2	20
700046677760671	1	Not seen
700076777760771	1	92
700777747413771	1	129
703377400001771	2	21
703777740003171	1	25
737417606060731	1	805
737777607030771	1	Not seen
737777677670371	1	Not seen
757777606060771	1	Not seen
757777777413731	2	806
770077606060731	2	807
773777606060731	1	Not seen
774777777760771	1	801
777377607760771	1	81
777417606060731	8	184
777737777760771	4	37
777747606060771	1	808
777757606060731	1	82
777774606060731	1	Not seen
777775606060771	3	809
777775770020771	1	Not seen
777777414020731	1	Not seen
777777602060771	1	810
777777604060731	7	811
777777606060571	3	812
777777606060631	4	813
777777606060671	3	814
777777606060731	7	815
777777606060771	68	59
777777607760731	1	60
777777607760771	23	42
777777674063771	1	Not seen
777777676060771	1	Not seen
777777677760771	5	291
777777757760771	1	44
777777757763771	1	Not seen
777777773413731	1	Not seen
777777774020731	2	62
777777774020771	1	47
777777774060771	1	Not seen
777777777413771	1	236
777777777420731	1	817
777777777473771	1	Not seen
777777777760671	1	245
777777777760771	25	53

[a] Octal codes determined using Dale *et. al* (2001) protocol.

[b] Obtained from the Pasteur Institute, Guadeloupe.

Table 4.3 Spoligotype clades observed in Harare, Zimbabwe and London, England



Spoligotype	n of total spoligotypes in Harare	n of total spoligotypes in London	RR (95%CI)
Afri1	0	3	-
Afri2	0	4	-
Afri3	0	0	-
BCG	0	0	-
Beijing	4	99	6.59 (2.45-17.70)
CAS1	0	120	-
CAS2	0	18	-
EAI1	0	34	-
EAI2	0	8	-
EAI3	0	43	-
EAI4	0	1	-
EAI5	1	2	0.53 (0.05-5.84)
H37Rv	0	2	-
Haarlem1	1	20	5.32 (0.72-39.44)
Haarlem2	0	6	-
Haarlem3	0	78	-
LAM1	2	12	1.60 (0.36-7.08)
LAM10	0	29	-
LAM2	0	3	-
LAM3	0	0	-
LAM4	1	5	1.33 (0.16-11.33)
LAM5	0	3	-
LAM6	0	11	-
LAM7	0	0	-
LAM8	0	0	-
LAM9(LAM1	23	26	0.30 (0.18-0.52)
M.cannetti	0	0	-
M.microti	0	0	-
S(LAM2)	0	9	-
T1	25	147	1.57 (1.05-2.33)
T2	0	27	-
T3	0	20	-
T4	0	2	-
X1	1	21	5.60 (0.76-41.42)
X2	0	38	-
X3	1	13	3.46 (0.46-26.30)

Spoligotyping clades described by Filliol *et al.* 2002. [a] LAM 9 formerly LAM. [b] S formerly LAM2. RR – rates ratio, CI – confidence interval

4.3.6 Spoligotyping families observed in Harare, Zimbabwe

A comparison of the spoligotypes observed in this study to the published spoligotype clades (Filliol *et al.* 2002) was performed. Of the 214 isolates with a spoligotype only 59 (27.6%) belonged to a recognised spoligotype family. Out of the 36 spoligotype clades described by Filliol *et al.* only 9 were observed (Table 4.3). The majority of isolates belonging to a spoligotype family (81.3%) were seen in just 2 families, the T1 and the (Latin American-Mediterranean) LAM 9 families (11.8% and 10.8% respectively of the total number of isolates). The next largest family, the Beijing family, was seen in just 1.9% of the total isolates.

4.3.7 Spoligotyping families observed in London, UK

All spoligotypes from cases isolated in the London region between 1st Jan 1998 and 31st March 1999 identified in the previous study (Chapter 3) were analysed to determine the demographics of spoligotype families observed in this region and in comparison to Harare. A total of 27 *M.tb* spoligotype families were observed in the London region (Table 4.3). From the total cases isolated during the study 46.0% had spoligotypes belonging to a recognised family. The most frequently seen spoligotype strains were T1 (18.3% of the total number of isolates) followed by (Central Asian family) CAS 1 (14.9%). Overall 99 Beijing/W strains were identified which comprised 12.3% of the total. The Haarlem 3 strain was also prevalent and found in 9.7% of cases. The spoligotype signature of the *M.tb* laboratory strain H37Rv was observed in two cases. In addition, seven spoligotypes characteristic of *M. africanum* were observed.

A statistical evaluation of the prevalence of different strain families in these two settings was performed using a rates ratio (RR). As the majority of strain families were not seen in Harare the comparison was limited to 9 families of which 3 showed a statistically significant difference. The Beijing and T1 strains were more prevalent in London than in Harare, RR 6.56, 95% CI (2.45-17.70) and RR 1.57, 95% CI (1.05-2.33) respectively, whereas the LAM 9 strain was less prevalent in London than in Harare, RR 0.30, 95% CI (0.18-0.52).

4.4 Discussion

4.4.1 Recent transmission versus reactivation of disease

The transmission of TB in an epidemiologically characterised population in Harare, Zimbabwe was assessed using two molecular techniques; spoligotyping and VNTR-ETR. Clusters of strains are thought to reflect recent transmission (Small *et al.* 1994, Alland *et al.* 1994, van Deutekom *et al.* 1997). Almost three-quarters of TB cases in Harare, Zimbabwe were clustered and so could potentially represent the rate of recent transmission. Reactivation of latent infection therefore appears to be low and accounts for very few TB cases in Harare. It is possible that had the collection of isolates been typed additionally using RFLP-IS*6110*, clustering rates would have been reduced.

Nevertheless, the level of clustering seen in this study was substantially higher than in countries with a low prevalence of TB which have been estimated at; 28% in France (Torrea *et al.* 1996), 38% in Seville, Spain (Safi *et al.* 1997), 38% in New York city, US (Alland *et al.* 1994), 40% in San Francisco, US (Small *et al.* 1994) and 46% in Amsterdam, The Netherlands (van Deutekom *et al.* 1997). Fewer molecular epidemiological studies have been carried out in countries

where rates of TB and HIV are endemic. Transmission has been estimated at 32% in San Paulo, Brazil where about half of cases were HIV positive (Ferrazoli *et al.* 2000), 33% in Guadeloupe (Sola *et al.* 1997), 42% in Botswana where 69% of cases were HIV positive (Lockman *et al.* 2001), 45%-50% in South Africa (Godfrey- Faussett *et al.* 2000, Wilkinson *et al.* 1997) and 72% in Northern Malawi (Glynn *et al.* 2005a). A small TB study in Harare conducted in 1995 found a clustering rate of 53.6% based on spoligotyping, (Heyderman *et al.* 1998) which was lower then observed in this study, however only 28 patients were assessed and so this is probably not a true representation.

The high rates of clustering seen in this study may reflect the high prevalence of HIV infection in this population. Almost 75% of cases were co-infected with HIV. HIV infection increases the susceptibility to exogenous TB infection and the rapid onset of TB after infection which in turn would lead to a higher prevalence of infectious cases over a shorter period of time. A similar rate of clustering was observed in Northern Malawi where clustering in older adults was associated with patients who were HIV positive suggesting that HIV has a greater impact on disease caused by recent transmission then that caused by reactivation (Glynn *et al.* 2005a).

Overcrowding in the high-density housing in Harare may also have lead to increased transmission rates. The study group was also relatively young which has previously been identified as a risk factor for clustering (e.g. Maguire *et al.* 2002). Other previously described risk factors in industrialised countries include; HIV infection, being a member of an ethnic minority, drug resistance,

homelessness, alcohol and drug abuse (Alland *et al.* 1994, Small *et al.* 1994, van Deutekom *et al.* 1997). None of these factors were found to be significant in this study, however it can be difficult to compare risk factors in different populations as definitions of certain characteristics vary. For instance the interpretation of 'homelessness' can vary depending on the country examined.

The short duration of the study (six months) and the fact that less than half of cases had molecular data may mean that links were missed. A comparison of patient characteristics in those with and without molecular data were similar, indicating that the results of this study are unlikely to be biased by availability of molecular data. Patients were also recruited from the main TB referral hospital which sees the majority of TB cases in Harare, therefore patients enrolled are likely to be representative of all the TB cases in the city. By only sampling smear positive cases a snapshot of the most efficient sources of transmission were identified, however smear negative, culture positive cases can still account for a sizable amount of TB transmission (Behr *et al.* 1999). Together these factors may have reduced the number of potential molecular epidemiological links and in turn contributed to an underestimation of recent transmission.

Strains with identical fingerprints in rural areas have been attributed to the simultaneous reactivation of infection from a common source at some time in the past (Braden *et al.* 1997), however no clustering was observed in rural home districts suggesting that these cases were due to reactivation of different strains.

Clearly clustered cases represent TB transmission between individuals at some point. Thereafter, the interpretation of clustering is debatable as identical fingerprints are thought to represent clusters of related cases which are more likely to be due to recent transmission rather then simultaneous reactivation of a latent infection. However a study of *M.tb* isolates in Malaysia found that clustering increased with patient age and identical fingerprints were seen over a wide area suggesting that reactivation may have contributed to clustering (Dale *et al.* 1999). In this Zimbabwean population reactivation is unlikely to have contributed greatly as the majority of cases were of a young age.

Although this study identified high rates of clustering, if the index case is presumed to be a member of each cluster and is therefore removed (i.e. n-1) (Small *et al.* 1994, Glynn *et al.* 2005a) the level of clustering is reduced from 84.1%, 89.4%, 73.0% to 74.8%, 83.3%, 60.3% for spoligotyping, VNTR-ETR and combined respectively. This method is perhaps a more accurate measure of the level of recent transmission however it does not take into account any cases that are infected just before the study period and subsequently mislabelled as a reactivated case.

Based on a combination of spoligotyping and VNTR-ETR just over a quarter of TB cases were unique and may represent reactivation of a latent infection. Due to the high prevalence of TB in Africa many children are likely to be exposed to and infected with TB before they reach adulthood, therefore reactivation of an infection acquired as a child could account for a large proportion of TB cases in Africa. A study of pulmonary TB cases in Botswana found half of all cases were

not linked and may have been due to reactivation of a latent infection (Lockman *et al.* 2001), however in this study recent transmission and not reactivation of disease appears to be the major cause of TB in Harare.

The typing method(s) chosen for molecular epidemiological studies is important as the stability of markers can vary, therefore clustered cases may not always represent recent transmission (Glynn *et al.* 1999). RFLP-IS*6110* is the gold standard for strain typing *M.tb*. Unfortunately RFLP-IS*6110* could not be carried out in this study due to the high level of heavy contamination of cultures and also due to sparse growth. As demonstrated in chapter 3, VNTR-MIRU is more discriminatory than spoligotyping and appears to have a discriminatory power similar to that of RFLP-IS*6110* analysis. The application of RFLP and VNTR-MIRU to strains in this study may sub-divide clusters. Unfortunately due to the low amounts of DNA remaining after spoligotyping and VNTR-ETR typing, VNTR-MIRU could not be performed.

PCR techniques are valuable in areas where the culture of tuberculosis is difficult as relatively little equipment and expertise is required. Spoligotyping and VNTR-ETR are suited to this kind of environment as they can be applied directly to smear-positive sputum. However, as spoligotyping and VNTR-ETR are not as discriminatory as RFLP-IS*6110* (Filliol *et al.* 2000 and Goyal *et al.* 1997) focus should be turned to PCR tools with a higher level of discrimination such as VNTR-MIRU.

4.4.2 Comparison of spoligotypes observed with an international spoligotype database

A comparison with an international spoligotype database showed that the three most prevalent spoligotypes were all ubiquitous in Africa. Type 59, the largest spoligotype cluster, is found almost exclusively in Africa and had previously been seen in Harare (Heyderman *et al.* 1998). The second and third largest clusters are also ubiquitous but were not observed in the Heyderman *et al.* study, although this study only contained 28 cases. Spoligotypes numbered between 804 and 815 which had not been seen before were similar to type 59 and may be considered evolutionary descendants. Type 62 is likely to be an emergent clone from type 47 which is from the Haarlem family of strains. Type 21 is found in East Africa and Europe, and therefore may be due to immigration from these countries.

4.4.3 Spoligotyping families observed in Harare and London

Worldwide analysis of *M.tb* spoligotypes has shown that the distribution of types can be confined to certain areas while others are widespread globally. Spoligotyping has been used as a phylogenetic marker to infer evolutionary relationships between these types or families of strains (Sola *et al.* 2001a, Sola *et al.* 2001b, Filliol *et al.* 2002, Filliol *et al.* 2003). The naming of these families has largely been dependant on the area in which they were first recognised or most prevalent, for example, the Beijing type (van Soolingen *et al.* 1995), Haarlem type (Kremer *et al.* 1999,) the East African-Indian and Latin American-Mediterranean families (Sola *et al.* 2001a).

The spoligotypes observed in two capital cities, one in a high incidence area Harare, Zimbabwe and the other in a relatively low incidence area London, UK were compared. In Harare, the majority of strains belonging to spoligotype clades (81.3%) were part of just two main families, the T and the LAM family whereas in London 13 large families were observed (those containing 20 or more isolates) which comprised 87.5% of isolates in a spoligotype clade, probably reflecting the greater international diversity of the population in London.

The comparison of spoligotype families observed in Harare and London found that the T1 strain was ubiquitous in both populations. The T family is thought to be a relatively old genotype (Filliol *et al.* 2002) and the T1 strain is found worldwide (Filliol *et al.* 2003), therefore it is not surprising that this strain was frequently seen in these two contrasting settings. The only member of the T family seen in Harare was the T1 strain, whereas in London all four T family strains were observed suggesting that some T strains may be more successful in certain regions. However it is possible that cases in Harare that could not be spoligotyped may have been members of the other T family strains.

The second most frequency observed spoligotype strain in Harare (LAM 9) which was seen in 10.8% of cases was also observed in 3.2% of cases in London; the second most frequently seen strain in London (CAS 1) was not observed in Harare. The CAS family are newly described and are generally found in middle-eastern Asia (Filliol *et al.* 2002) whereas the LAM family is believed to have originated in either the Americas or in the Mediterranean region and later spread to the Americas through migration of infected cattle and humans (Sola *et al.*

2001a). The LAM family belongs to genetic group 2 and they are therefore thought to be evolutionary intermediate strains (Streevatsan *et al.* 1997) and are also thought to share a common ancestor with the Haarlem family (Sola *et al.* 2001a). The three Haarlem strains comprising the Haarlem family were all seen in London, particularly Haarlem 3, which was observed in 9.7% of cases. In Harare, however, only Haarlem 1 was observed (0.5% of strains).

Surprisingly the (East African-Indian) EAI family, which was thought to have originated in Africa or Asia (Sola *et al.* 2001a), was not over-represented in Harare. In fact only one member of this family, EAI-5, was observed (0.5% of cases). In contrast, all members of the EAI family were seen in London isolates and comprised 10.9% of the total isolates. Moreover, a spoligotype with the EAI pattern was isolated from a 15th century skeletal sample of *M.tb* DNA found in a UK village and so spreads doubt on the hypothesised origin of this strain (Mays *et al.* 2001).

M. africanum strains are commonly associated with TB patients in Africa (Niemann *et al.* 2002) however no spoligotype patterns from this species were observed in Zimbabwean isolates, conversely 7 *M. Africanum* strains were observed in London isolates, although this species has previously been seen in South-East of England (Grange and Yates 1989).

The Beijing strain family was first identified in the Beijing area of China, where a high proportion of strains examined belonged to a genetically closely related group (van Soolingen *et al.* 1995). This strain has since disseminated globally

and is particularly prevalent in former Soviet Union countries (Drobniewski *et al.* 2002, Bifani *et al.* 2002). Beijing strains are easily recognisable by their distinctive spoligotype pattern which only contain spacers 35 to 43 (Fig 3.3). The Beijing strain was observed in both settings although at a higher proportion in London (12.3% versus 1.9% in Harare) and demonstrates that the Beijing strain has continued to spread globally. The Beijing strain family are believed to be intrinsically more virulent then non-Beijing strains causing accelerated onset and more severe disease (Lopez *et al.* 2003). A related strain, the W strain, was the cause of a multi-drug resistant outbreak in New York City and has since spread across the USA. This strain shares the same spoligotype pattern as Beijing (Bifani *et al.* 1996, Agerton *et al.* 1999). None of the Beijing/W strains isolated in Zimbabwe were drug resistant and only 2/99 (2.0%) of strains isolated in London were resistant to one or more anti-tuberculous drugs. Relatively little is known about the effect of Beijing/W strains in Africa (Glynn *et al.* 2002), however, drug-resistant cases have been reported in Cape Town and Nairobi (van Rie *et al.* 1999, Githui *et al.* 2004) but a seven year study in northern Malawi found all Beijing strains were fully drug sensitive (Glynn *et al.* 2005b).

Identifying which clades/strain families are present in various populations is not only important in terms of monitoring the global spread of TB but also in identifying which strains are successful within a population. Successful strains may harbour particular resistance patterns, such as the W-strain which has been linked to drug resistance, and therefore the spoligotypes of these strains can act as beacons for TB control action. In population-based studies these spoligotype

signatures can be used to screen isolates allowing the identification of potentially related cases for which public health action can be targeted.

Note:

This study was part of collaboration between the Department of HIV/GUM, King's College Hospital, UK, the HPA-Mycobacterium Reference Unit, UK, the University of Zimbabwe and Harare City Health Department, Zimbabwe and the Institute Pasteur, Guadeloupe. Andrea Gibson performed the molecular typing, except for some of the spoligotyping which was carried out by Dianie Lamprecht, and molecular analysis. The epidemiological analysis was performed mainly by Professor Philippa Easterbrook and Dr. Shahed Murad.

CHAPTER 5.0 Molecular epidemiology of *Mycobacterium bovis* disease in humans in the UK

5.1 Introduction

Mycobacterium bovis, the causative agent of bovine tuberculosis, has a wide host range infecting many domestic and wild animals. *M. bovis* infection in humans is relatively rare in developed countries, but it remains of public health importance in developing countries (Cosivi *et al.* 1998). Approximately 1% of bacteriologically proven clinical tuberculosis (TB) is attributed to *M. bovis* in the UK (http://www.hpa. org.uk/infections/topics_az/tb/mbovis). The resurgence of bovine TB in cattle in the UK has raised concerns that transmission from cattle to humans might be a serious public health issue.

M. bovis was once a major source of TB in humans in the UK but was almost eradicated after the introduction of control measures to reduce bovine tuberculosis in cattle together with the pasteurisation of milk for human consumption. The majority of human cases of *M. bovis* in the 1980s and early 1990s in the UK were due to reactivation of latent *M. bovis* infections which were acquired pre-milk pasteurisation or due to cases who were infected abroad and have subsequently returned or migrated to the UK (Hardie and Watson 1992). The increase in bovine TB in cattle, particularly in the southwest of England, has caused concern about the subsequent transmission to humans. Many animals such as badgers, foxes, ferrets and deer are believed to act as vectors for transmission to livestock (Aranaz *et al.* 1996; Clifton-Hadley *et al.* 1995; Delahay *et al.* 2001) and some have been associated with transmission to

humans (Dalovisio *et al.* 1992; Michalak *et al.* 1998; van Soolingen *et al.* 1994). It is therefore important to rapidly identify where rates of *M. bovis* in cattle are high and pose a potential risk of transmission to humans.

Advances in molecular typing have enhanced our knowledge of *M. bovis* dissemination. Although RFLP-IS*6110* analysis is considered the gold standard for typing *M.tb, M. bovis* isolates from cattle usually have only a single copy of IS*6110* (Cousins *et al.* 1998) and therefore alternative techniques such as spoligotyping and variable number tandem repeats (VNTR) have been successfully used in discriminating between strains of *M. bovis* in animals (Aranaz *et al.* 1996; Cousins *et al.* 1998; Frothingham and Meeker-O'Connell 1998; Haddad *et al.* 2001; Kamerbeek *et al.* 1997; Roring *et al.* 2002).

5.2 Study design

All available viable *M. bovis* isolates (50 isolates) from humans diagnosed in the UK between 1997 and 2001 were identified; 40 from the HPA Mycobacterium Reference Unit (MRU), London and 10 from the Scottish Mycobacteria Reference Laboratory (MRL), Edinburgh. A crude DNA extract (Materials and Methods 2.2.1) of each isolate was spoligotyped and VNTR-ETR typed (Materials and Methods 2.4 and 2.5.1). One strain with a spoligotype not typical of *M. bovis* was sent for deletion analysis at the Veterinary Laboratory Agency (VLA). The discriminatory power of each typing method was determined using the Hunter-Gaston index (HGI) (Material and Methods 2.7). Epidemiological information was obtained from laboratory records at the MRU and MRL and

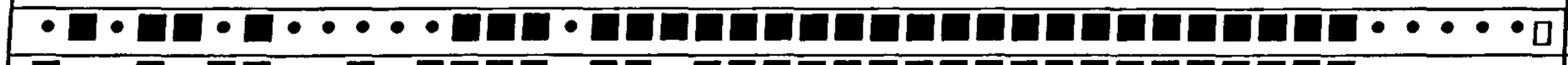
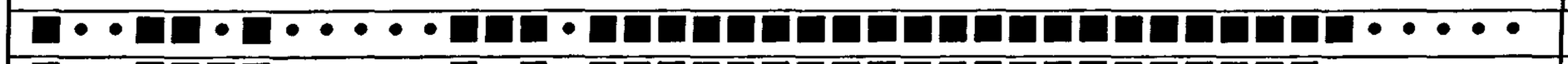
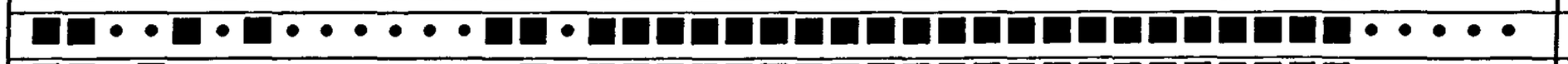
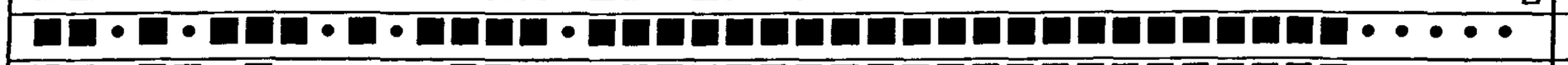
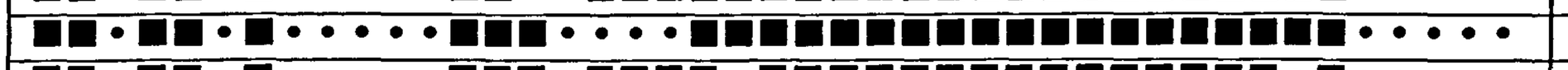
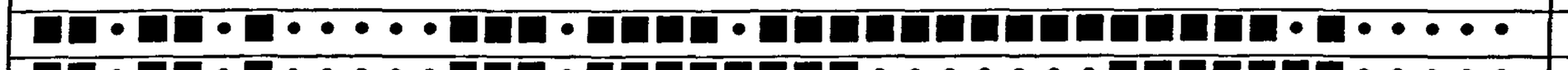
from surveillance data held at the Health Protection Agency (HPA) Communicable Disease Surveillance Centre (CDSC).

5.3 Results

Spoligotyping of the 50 human *M. bovis* isolates produced 25 distinct spoligotypes with a HGI of 0.892. Thirty-two isolates were divided into 7 clusters A to G (Table 5.1). The spoligotypes were compared with *M. bovis* spoligotypes from a bank of over 15,000 cattle isolates collected from all over the UK held at the Veterinary Laboratory Agency (VLA) dating between 1987-2002. The largest cluster of human *M. bovis* isolates (15 isolates, 30%) had been seen in cattle before and was sequentially numbered type 9 (international type SB140 – http://www.Mbovis.org) at the VLA. Type 9 is the most frequently seen spoligotype of *M. bovis* (over 30 % of all isolates have this spoligotype) isolated from cattle and has a wide geographical range in the UK (Durr *et al.* 2000).

Human type 9 isolates were seen across the UK, suggesting that transmission between cattle and humans might occur. Interestingly, 15 of the human *M. bovis* spoligotypes had not been seen at the VLA in cattle isolates. When compared to the international spoligotype database, only 2 of the 15 types were recognised. The first was isolated in Argentina and the second in Australia, the remaining 13 spoligotypes were all unique to the UK. In the majority of these cases it is likely that disease was due to reactivation of a past infection that had been acquired prior to milk pasteurisation, rather then primary infection, because 11/13 (84.6%) cases were over the age of 50. The two cases under the age of 50 were part of an epidemiologically conformed case of transmission described on page 103.

Table 5.1 Spoligotyping patterns for human *M. bovis* isolates in the UK

Human *M. bovis* spoligotypes	VLA* Type	Internation al Type**	n of isolates	Cluster[Ŧ]
	31	SB0155	1	
	22	SB0673	1	
	Ns[b]	Ns	1	
	Ns	Ns	1	
	Ns	Ns	1	
	Ns	Ns	1	
	35	SB0134	1	
	Ns	Ns	1	
	20	SB0145	1	
	6	SB0267	1	
	Ns	SB0131	2	A
	9[c]	SB0269	2	B
	Ns	SB0144	1	
	Ns	Ns	1	
	Ns	Ns[d]	2	C
	17	SB0263	1	
	10	SB0272	3	D
	13	SB0273	1	
	9[a]	SB0140	15	E
	Ns	Ns	1	
	Ns	Ns	1	
	21	SB0130	6	F
	Ns	Ns	1	
	BCG	SB0120	2	G
	Ns	Ns	1	

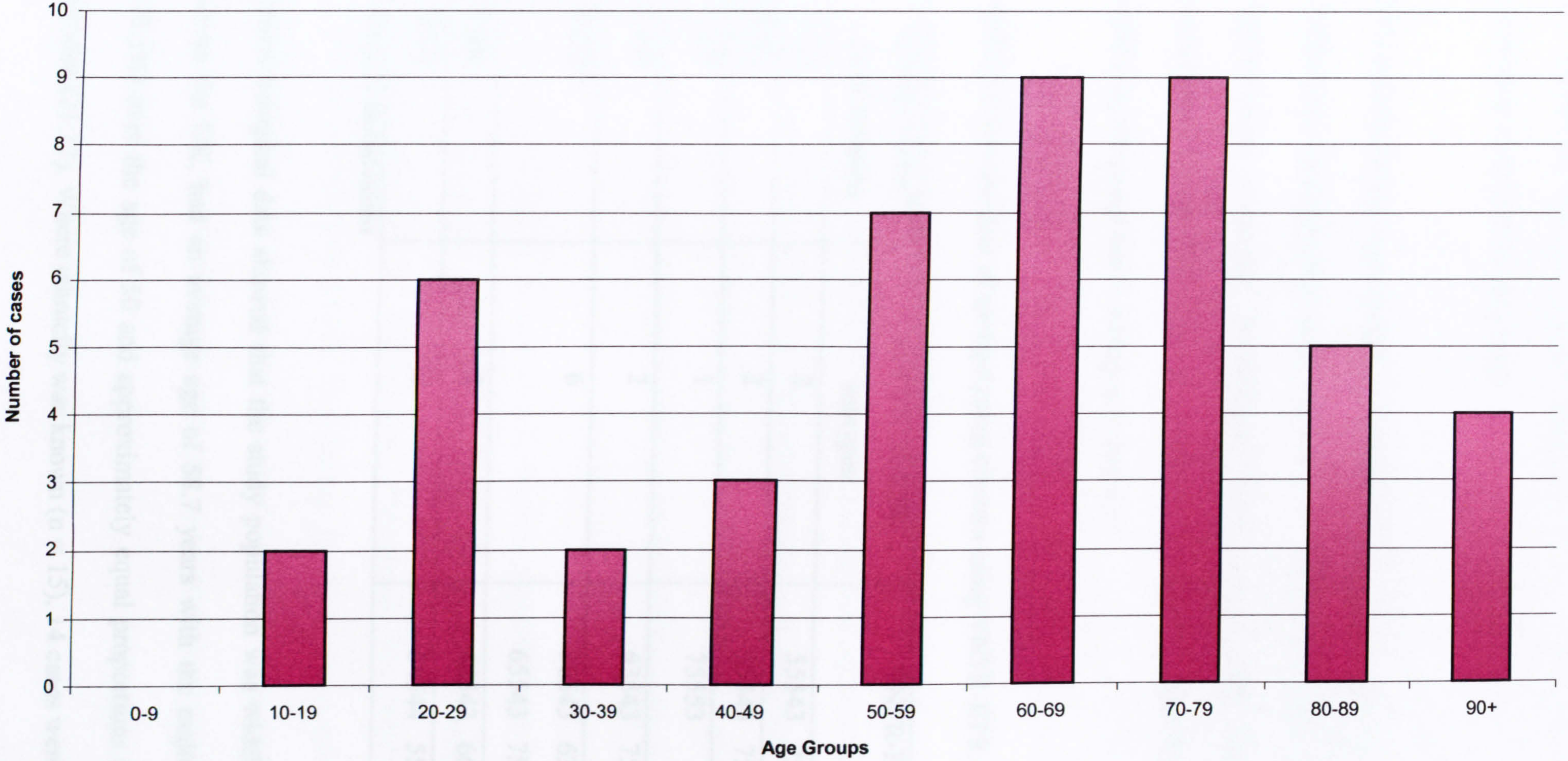
* Comparison of human *M. bovis* spoligotypes with a bank of *M. bovis* spoligotypes seen in UK cattle

**International spoligotype number (www.Mbovis.org)

[Ŧ] Arbitrarily labelled clusters

[a] Most predominant spoligotype (Type 9), [b] Spoligotype not seen in UK cattle before, [c] Identical to Type 9 except for absence of spacer 15, [d] Farm outbreak strain

Figure 5.1 Number of *M. bovis* cases by age group



Excluding these two cases the age range was 51 to 96 years with a median age of 67. Therefore, these 11 unique spoligotypes may reflect *M. bovis* strains circulating in the UK 50 years ago.

VNTR-ETR typing alone produced 18 different profiles and had a HGI of 0.844. Combining spoligotyping with VNTR-ETR vastly improved the level of discrimination producing 34 different types with a high HGI of 0.958. Furthermore VNTR-ETR was very useful in subdividing type 9 spoligotypes, separating the group into 6 subtypes (Table 5.2).

Table 5.2 Sub-division of spoligotyping clusters using VNTR-ETR

Spoligo cluster* (n of isolates)	n of VNTR-ETR subtypes	VNTR-ETR profiles
A (2)	2	55543 75543
B (2)	2	55543 75543
C (2)	1	75553
D (3)	2	63543 75543
E (15)	6	56543 63543 65542 65543 75543 75553
F (6)	2	65543 66543
G (2)	2	54544 55343

*Arbitrarily labelled clusters

Epidemiological data showed that the study population was widely distributed across the UK, had an average age of 58.7 years with the majority of cases (72.3%) over the age of 50 and approximately equal proportions of males and females (21:17). Where ethnicity was known (n = 15), 14 cases were white and 1

was of Black-African origin who was originally from Nigeria but had lived in the UK since 1996. This person had a unique spoligotype therefore it is possible that she was infected in Nigeria before arriving in the UK. Of interest, 59% (13/22) of cases had some contact with a farm, ranging from a part-time job milking cows, living on a dairy farm as a child, to being a farmer (now retired). One spoligotype cluster represents an outbreak on a farm in Gloucester which was established by this work. Two siblings, (a 20 year old male and a 17 year female) living on their parents' farm became infected with *M. bovis*. The brother occasionally helped his father on the farm by restraining the cattle and would often be sprayed with nasal mucus. Cattle infected with *M. bovis* of the same spoligotype had been detected on the farm in previous years. Transmission from cattle to human is thought to have occurred by inhaling infected aerosols from cattle. The brother is thought to have subsequently infected his sister, as she had no contact with the cattle but was also diabetic and pregnant i.e. immuno-compromised. This is thought to be the first documented case of animal to human transmission since the resurgence of bovine TB in the UK (Smith *et al.* 2004).

M. bovis spoligotypes do not usually contain spacers 39-43, however, one spoligotype from the panel contained spacers 40 to 43, which are more commonly seen in *M. tuberculosis.* Phenotypic and biochemical tests demonstrated that this isolate had typical *M. bovis* characteristics; it was microaerophilic, TCH (thiophen-2-carboxylic acid hydrazide) negative, pyrazinamide resistant and grew better on pyruvate than glycerol Lowenstein-Jensen slopes.

Deletion analysis was performed (at the VLA) to ascertain the identity of this strain. The strain contained RD 4, 12, and 13 but lacked RD 7, 8, 9 and 10 (Brosch *et al.* 2002) indicating that this strain was actually *M. africanum* and not *M. bovis*.

5.3.1 VNTR-MIRU compared to VNTR-ETR and spoligotyping in an *M. bovis* population

Based on the analytical strategy described in chapter 3 each of the *M. bovis* isolates in this panel were VNTR-MIRU typed using 15-loci. Of the 49 *M. bovis* isolates (excluding the isolate confirmed as *M. africanum*), 45 VNTR-MIRU types were observed which included 4 clusters each containing 2 isolates. This dramatic decrease in the number of clusters indicates VNTR-MIRU is more discriminatory than either spoligotyping or VNTR-ETR alone or in combination (Table 5.3).

Table 5.3 Comparison of the discriminatory power of each typing technique

Technique	n of clusters	Total n of types	n of clustered isolates (%)	n of unique types	HGI
Spoligotyping	7	24	32	17	0.892
VNTR-ETR	7	17	39	10	0.844
VNTR-MIRU- 12	8	29	28	21	0.952
VNTR-MIRU- 15	4	45	8	41	0.997
Spoligotyping + VNTR-ETR	5	33	20	29	0.958
Spoligotyping + VNTR-MIRU-12	4	45	8	41	0.997
Spoligotyping + VNTR-MIRU-15	2	47	4	45	0.999

VNTR-MIRU-15 gave the best discrimination of this bovine population followed by VNTR-MIRU-12, spoligotyping and VNTR-ETR (0.997. 0.952, 0.892, 0.844 respectively). When combined, spoligotyping and VNTR-MIRU-15 had the best discriminatory power (0.999). VNTR-MIRU-15 alone had a discriminatory power almost equal to that of both techniques combined however it produced more clustered isolates. VNTR-MIRU-15 was useful in sub-dividing type 9 spoligotypes: 15 isolates were sub-divided into 14 different subtypes suggesting VNTR-MIRU-15 would be more useful for epidemiological studies than spoligotyping or VNTR-ETR.

Apart from one exception, there was no epidemiological evidence to support the four clusters linked by VNTR-MIRU-15. The epidemiologically linked cluster contained two siblings involved in the farm outbreak of *M. bovis* (described on page 103). Of the three remaining VNTR-MIRU-15 clusters, only one contained isolates with the same spoligotype (type 9). The remaining two clusters were, in each case, composed of one isolate with a known pattern and one with a pattern not previously seen in the UK.

5.4 Discussion

Although the number of bacteriological conformed cases of TB caused by *M. bovis* remains small in the UK, it is important to continue to monitor bovine TB in humans. This is particular important in those who are at high risk of primary infection such as agricultural and abattoir workers. The increasing popularity of holidays in parts of the world where *M. bovis* disease in animals is common (transmission can occur through the consumption of unpasteurised milk and

cheese) provides another source of infection. Within the UK, high rates of bovine TB in cattle could potentially increase the risk of transmission to humans, therefore a strict test and slaughter policy, systematic inspection of meat carcasses and the monitoring of cattle movement are essential. In addition the adequate pasteurisation of milk for human consumption remains a corner stone of public health control.

A combination of spoligotyping and VNTR-ETR is an efficient discriminatory tool for the molecular surveillance of *M. bovis* and also addresses the problem of analysing isolates with single copies of IS*6110*. The most frequently observed spoligotype in humans in the UK was type 9 which is also the most frequently seen in cattle. This spoligotype was sub-divided by VNTR-ETR into 6 different types. The geographical distribution of VNTR-ETRs in *M. bovis* isolates with the most frequently observed spoligotype patterns (type 9 and type 17) in Great Britain, suggests that *M. bovis* has undergone a series of clonal expansions (Smith *et al.* 2003). Only one type 17 spoligotype was seen in this study, whereas type 9 spoligotypes were observed in 30% of isolates. The type 9 strains included VNTR-ETR patterns 6554 and 7554 (ETR A to D) of which the former is the most frequently seen in isolates with type 9 spoligotypes and latter is thought to be the ancestral type due to its wide distribution (Smith *et al.* 2003).

The application of VNTR-MIRU has shown a high level of discrimination in this *M. bovis* population and was able to sub-divide the most prevalent spoligotype – type 9. However two VNTR-MIRU clusters were sub-divided by spoligotyping. As both techniques analyse different parts of the genome it is difficult to

decipher which is the more accurate. Spoligotyping (as seen with the type 9 isolates) tends to group isolates together which may not be linked. Conversely it is possible that the discriminatory power of VNTR-MIRU may be too high resulting in the break up of clusters that are linked. Evidence in this population, however, shows this is unlikely as the epidemiologically confirmed cluster was not sub-divided. Indeed in a study of *M.tb* isolates all RFLP-IS*6110* and epidemiologically linked isolates were clustered by VNTR-MIRU (Kwara *et al.* 2003). VNTR-MIRU also has the advantage of analysing multiple loci situated around the genome and therefore each locus is unlikely to be subjected to the same selection pressures i.e. VNTR-MIRU may reflect a fairer picture compared to spoligotyping which only analyses one locus.

VNTR-MIRU has proven to be a good technique for the analysis of *M. bovis* isolates. As a PCR-based technique, VNTR-MIRU is rapid and requires only small amounts of DNA. Furthermore, VNTR-MIRU-15 has a higher discriminatory power then many of the other commonly used PCR techniques therefore VNTR-MIRU-15 alone or in combination with spoligotyping should be the method of choice for studying *M. bovis* in human isolates. It is also highly likely that the VNTR-MIRU-15 technique would be far superior to the currently used spoligotyping or VNTR-ETR strategies employed for investigating animal or veterinary transmission.

CHAPTER 6.0 Phylogenetic analysis of *Mycobacterium tuberculosis* complex

6.1 Introduction

The advent of molecular techniques has dramatically increased our understanding of the transmission and dissemination of *M.tb* but despite these advances the evolution of this organism is still poorly understood.

Genotyping techniques based on neutral genetic variation such as multilocus sequence typing (MLST) have been successfully used to characterise bacterial populations (Maiden *et al.* 1998, Enright *et al.* 1998, Dingle *et al.* 2001, Gutacker *et al.* 2002, Baker *et al.* 2004). Although the *M.tb* genome is thought to be highly conserved (Sreevatsan *et al.* 1997 and Musser *et al.* 2000), sufficient neutral variation in the form of silent single nucleotide polymorphisms (sSNPs) was found in *M.tb* complex isolates from which a robust phylogenetic tree was constructed (Baker *et al.* 2004). The tree demonstrated that *M.tb* is clonal and can be sub-divided into 4 distinct lineages while *M. bovis* is closely related but found on a separate branch. The tree also accommodated other phylogenetic groupings including genetic groups 1 to 3 (Sreevatsan *et al.* 1997) and the *M.tb* specific deletion (TBD1) (Brosch *et al.* 2002). IS*6110* and spoligotyping families such as Beijing, Haarlem, Central Asian (CAS), Delhi and East African-Indian (EAI) families all appeared exclusively on different branches of the tree.

Variable Number Tandem Repeat - Mycobacterial Interspersed Repetitive Units (VNTR-MIRU) is a high-throughput technique which analyses the number of tandem repeats at loci distributed around the *M.tb* genome (Supply *et al.* 2001).

Studies have explored VNTR-MIRU as a tool for discriminating between strains and in comparison with other molecular techniques (Cowan *et al.* 2002, Hawkey *et al.* 2003, Kwara *et al.* 2003, Roring *et al.* 2004 and Supply *et al.* 2001) but there is less published data using VNTR-MIRU to study the evolution of the *M.tb* complex. VNTR using 5-loci and 12-loci MIRU have been used in combination with other techniques such as spoligotyping to investigate the evolution of *M.tb* complex (Ferdinand *et al.* 2004, Kam *et al.* 2005 and Sola *et al.* 2001a), but none have utilised 15-loci VNTR-MIRU to define evolutionary pathways characterised by sSNP analysis.

Based on the sSNP-defined phylogenetic tree (Baker *et al.* 2004) this study aimed to evaluate the use of VNTR-MIRU, commonly used in epidemiological investigations, as a rapid tool for the phylogenetic classification of *M.tb* and *M. bovis*. Validation of the VNTR-MIRU phylogenetic classification was achieved by characterising an additional panel of *M.tb* and *M. bovis* strains with a sSNP macroarray tool (the development of which is described in section 6.2.2) and examining the concordance with VNTR-MIRU.

6.2 Study design

6.2.1 Mycobacterial strains

A panel (panel 1- test population) of 312 *M.tb* clinical isolates collected in England and Wales between 1st January and 31st December 1998 included all isolates resistant to one or both of the first-line antituberculous drugs (isoniazid and rifampicin), 100 randomly chosen fully susceptible isolates and in addition four *M. bovis* isolates (Baker *et al.* 2004). These isolates were typed initially with

VNTR-MIRU-12 then supplemented with ETR A, B and C (VNTR-MIRU-15) (Materials and Methods 2.5.2). All members of this panel had previously been classified into one of five lineages according to the presence of characteristic sSNPs (Baker *et al.* 2004). The laboratory reference strain *M.tb* H37Rv (which had been characterised by sSNP analysis) was also VNTR-MIRU typed.

A second panel (panel 2) of 205 isolates which included 80 *M.tb* isolates identified in England and Wales in 1998 but not previously analysed by Baker *et al.* and 80 *M.tb* isolates identified in England and Wales in 2004 were analysed with VNTR-MIRU-15. Two different time windows, 1998 and 2004 were chosen to ensure that the results seen were reproducible and not unique to 1998, the year from which panel 1 strains were isolated. The *M.tb* isolates were all randomly chosen from 1998 and 2004. An additional 45 *M. bovis* isolates from a panel previously described in chapter 5 and by Gibson *et al.* (2004) were also analysed (the original four *M. bovis* from Baker *et al.* (2004) were drawn from this panel). Panel 2 was characterised with a macroarray developed in this study (see section 6.2.2) to define sSNPs seen in each of the five lineages. Sequencing these additional isolates to obtain sSNP data would have been very resource-intensive so a macroarray was developed for the rapid classification of these isolates into one of the five lineages.

Validation was performed blinded. Results were only compared once a lineage based on each technique had been determined. The macroarray was validated on a sub-panel of 46 previously genotyped isolates from panel 1.

6.2.2 Development of a macroarray to identify sSNPs

A combination of four sSNPs within three genes; *oxy*R-37, *katG*-87, *rpoB*-2464 and –3243 were able to define all five lineages (Baker *et al.* 2004).

Primers flanking each of the four sSNPs and probes targeting each sSNP site were designed. Both primers and probes were based on the DNA sequences of *M.tb* H37Rv and CDC1551 and checked for homology to other DNA sequences in the *M.tb* / *M. bovis* genome using BLAST (www.ncbi.nlm.nih.gov).

Wild-type (*wt*) (*M.tb* H37Rv sequence) and mutant-type (*mt*) (corresponding to the sSNPs defining each of the lineages described by Baker *et al.* 2004) probes of between 18 and 25 bases long were initially designed with the target sSNP, if possible, in a central position. If no hybridisation signal was observed the sSNP was moved to either end of the probe, but never within three bases of the end, until a hybridisation signal was observed. To optimise probe performance a poly-T tail of 10 bp was added.

The final macroarray method is described in Materials and Methods 2.8 and each of the probes tested in the following experiments is listed in Appendix 6. A series of experiments were performed to optimise the method for detecting the target sSNPs in this study and are described below.

Each probe was tested separately by blotting three dots onto a membrane using a pipette tip. Wild-type probes were tested using *M.tb* H37Rv (a lineage II isolate) and *mt* probes were tested with one clinical isolate corresponding to each of the

four remaining lineages. The four regions containing the sSNPs of interest (described in Material and Methods 2.8.1) were amplified in one isolate from each of the five lineages. After amplification 12 µl of each PCR product corresponding to one isolate from each of the five lineages was denatured using an equal amount of denaturing solution (400mM NaOH, 20mM EDTA) for 5 mins. The denatured solutions were incubated in separate tubes containing a membrane and hybridisation solution (5x SSPE, 0.5% SDS) for 1 hr at 60°C. Details of the hybridisation and detection protocol are detailed in Material and Methods 2.8.5 and 2.8.6 (N.B. Material and Methods 2.8.5 gives a final method. In the preliminary experiments 0.4X SSPE was used in the stringent wash buffer). Once a hybridisation signal was observed for each *wt* and *mt* probe the PCR products (i.e. four for each isolate) were combined to ascertain whether they would perform well together. To test this, a membrane containing all eight probes (4 *wt* and 4 *mt*) was designed (Material and Methods 2.8.3).

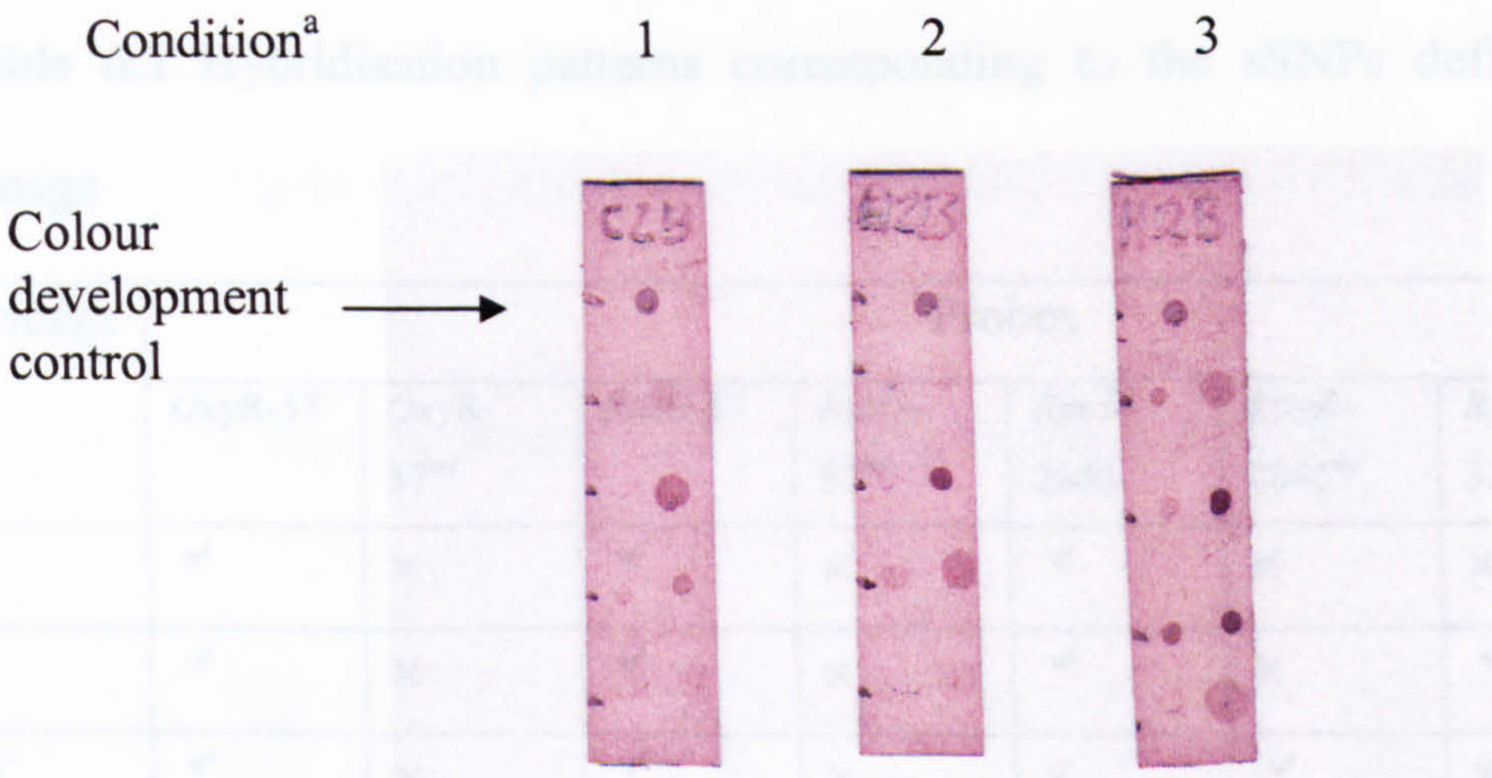
The four regions containing the sSNPs of interest (described in Material and Methods 2.8.1) were amplified in one isolate from each of the five lineages. After amplification 12 µl of each of the four PCR products corresponding to one isolate from each lineage was combined and incubated with a membrane containing all *wt* and *mt* probes. The hybridisation and detection procedure remained the same. No hybridisation signals were observed after completion of the experiment (results not shown). The concentration of NaOH was possibly too high for hybridisation to occur compared to the conditions when the probes were analysed separately (48 µl of NaOH versus 12 µl respectively). Solutions to this problem included using less DNA (hence reducing the amount of NaOH added)

or to denature the DNA using heat rather than chemicals. The experiment was repeated under three different conditions: 1) 2 µl of each PCR product plus 8 µl of denaturing solution, 2) 2 µl of each PCR product. Denature DNA at 100°C for 10 mins and 3) 12 µl of each PCR product. Denature DNA at 100°C for 10 mins. This time only *M.tb* H37Rv was tested as the aim was to determine which denaturising condition produced the best results. All three conditions gave hybridisation signals (Figure 6.1). Condition 3 gave the strongest hybridisation signal while chemical denaturation gave the weakest signal. Heat denaturation of 12 µl of each of the four regions of interest (containing the sSNPs) was used in all future experiments.

Using this method hybridisation of *M.tb* H37Rv DNA was observed at both the *wt* and *mt* probes at all sSNP sites. As *M.tb* H37Rv is a *wt* strain, hybridisation should only be observed at *wt* probes. The stringency of the washing conditions, e.g. by changing the temperature or the salt concentration at the hybridisation/washing phase, was altered to improve specificity.

A titration of salt concentrations was set up from 1X to 0.4X SSPE at 0.2X intervals then from 0.4X to 0.1X SSPE at 0.1X intervals as salt concentration could be controlled more accurately than temperature (the temperature of hybridisations ovens can vary). For this experiment clinical isolates from each of the four lineages (I, III, IV and *M. bovis*) were analysed in additional to *M.tb* H37Rv (a lineage II strain) in order to determine the optimum washing

Figure 6.1 The results of denaturing DNA under three different denaturing conditions



DNA amplification of the four gene regions (used to differentiate between the 5 lineages) was performed using *M.tb* H37Rv (i.e. the wild-type strain). The PCR products were combined and incubated with 1 of 3 membranes corresponding to 1 of the 3 different denaturing conditions (detailed below).

[a] Condition 1 - 2 µl of each PCR product plus 8 µl of denaturing solution. Condition 2 - 2 µl of each PCR product. Denature DNA at 100°C for 10 mins. Condition 3 -12 µl of each PCR product. Denature DNA at 100°C for 10 mins.

The single spot in the first row corresponds to the colour development control followed by four rows of *wt* (left hand side) and *mt* probes (right hand side) for each gene; *oxyR*-37, *katG*-87, *rpoB*-2646 and *rpoB*-3243 respectively. NB. The *wt* and *mt* probes for *rpoB* – 3243 were reversed in this experiment only.

conditions. In this experiment the expected hybridisation patterns for each lineage are described in table 6.1.

Table 6.1 Hybridisation patterns corresponding to the sSNPs defining each lineage

Lineage	Probes							
	*Oxy*R-37	*Oxy*R-37^{mt}	*KatG*-87	*KatG*-87^{mt}	*RpoB*-2646	*RpoB*-2646^{mt}	*RpoB*-3243	*RpoB*-3243^{mt}
I	✔	×	✔	×	✔	×	×	✔
II	✔	×	✔	×	✔	×	✔	×
III	✔	×	✔	×	×	✔	×	✔
IV	×	✔	✔	×	✔	×	×	✔
M. bovis	✔	×	×	✔	✔	×	×	✔

^{mt} – mutant type

✔ /× - positive/negative hybridisation signal

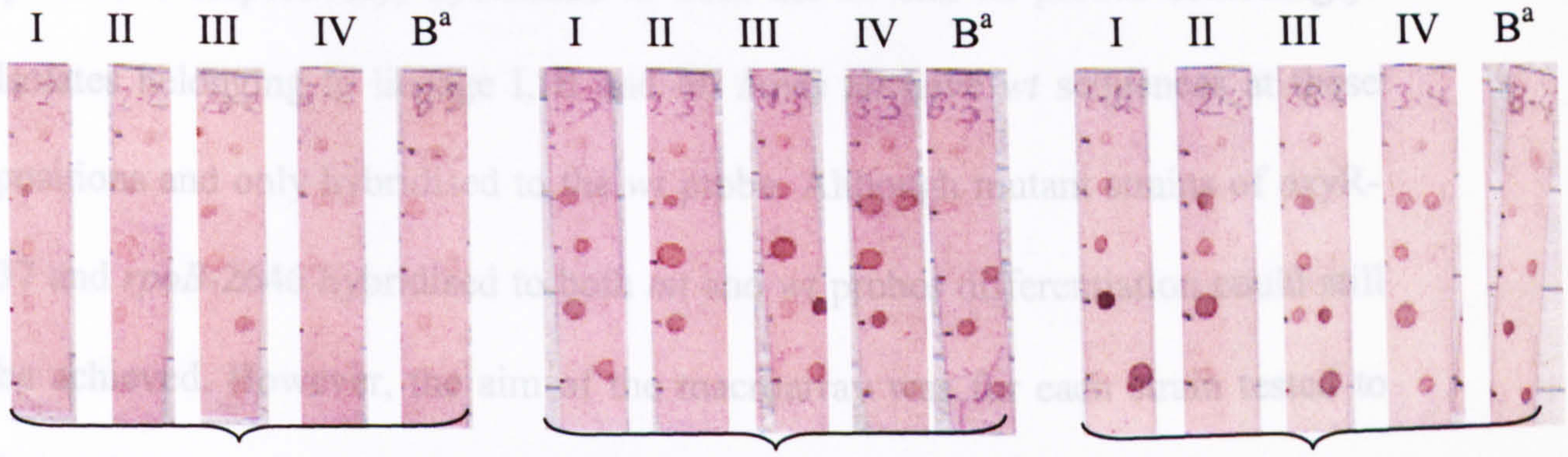
The results showed that 1X to 0.6X SSPE gave strong hybridisation signals but no differentiation between *wt* and *mt* probes (results not shown). A concentration of 0.4X and 0.3X SSPE gave strong hybridisation signals but variable discrimination between *wt* and *mt* probes whereas 0.2X and 0.1X SSPE gave better discrimination between *wt* and *mt* probes but fainter hybridisation signals (results not shown).

As some probes worked well at one salt concentration but not another further changes would have been counter-productive as increasing stringency to increase specificity would result in those probes with weaker hybridisation signals disappearing. For this reason all probes were re-designed. The probes were kept

as short as possible to try and increase specificity by making them less stable during washing and a longer poly-T tail (20bp) was added to increase adhesion.

Several batches of new probes were designed and tested at 0.1X, 0.2X, 0.3X and 0.4X SSPE. Figure 6.2 shows the results for 0.2X, 0.3X and 0.4X SSPE.

Figure 6.2 Stringency titration using 0.2X, 0.3X and 0.4X SSPE



[a] Corresponds to *M.tb* lineage I to IV. B = *M. bovis*

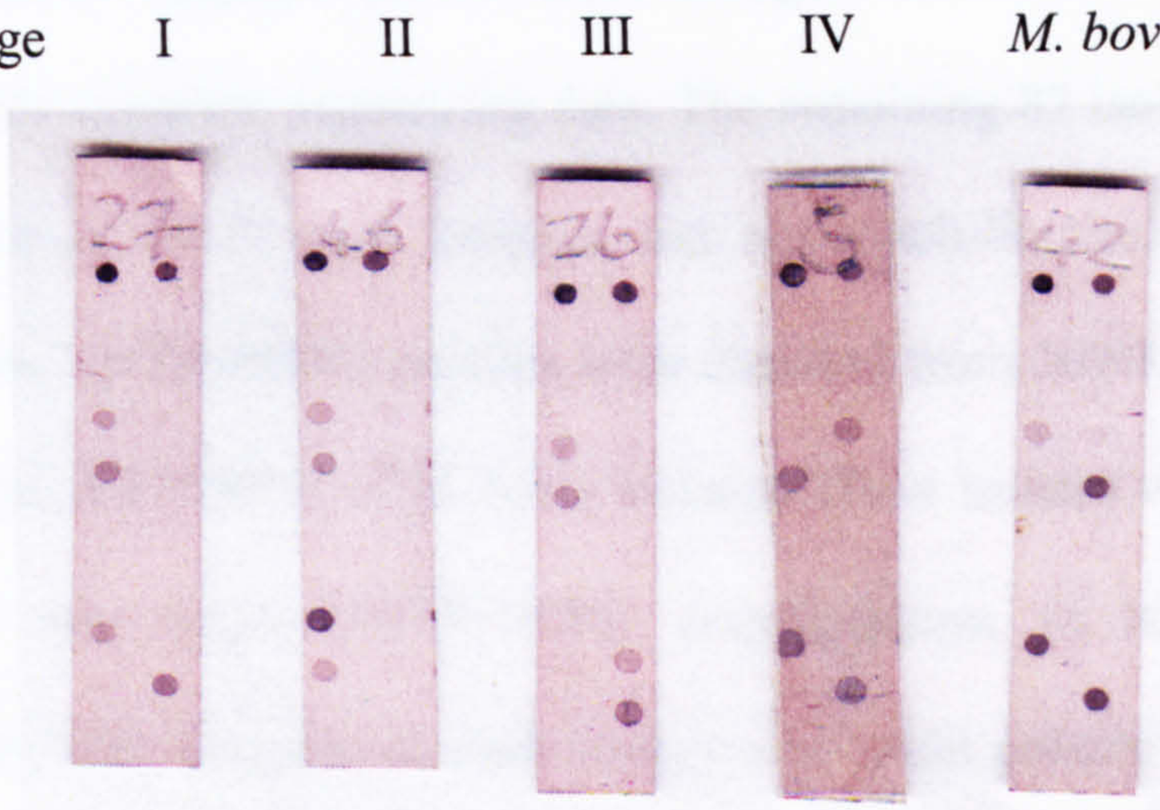
DNA amplification of the four gene regions was performed using an isolate from each of the 5 lineages. For each lineage the appropriate PCR products were combined and incubated with a membrane.

The single spot in the first row corresponds to the colour development control followed by four rows of *wt* (left hand side) and *mt* probes (right hand side) for each gene; *oxyR*-37, *katG*-87, *rpoB*-2646 and *rpoB*-3243 respectively. The expected hybridisation patterns are shown in table 6.1.

Salt concentrations of 0.3X and 0.4X SSPE gave good discrimination between *wt* and *mt* probes and the strongest intensity of hybridisation signal. In future experiments 0.3X SSPE was used.

At this stage an isolate belonging to any of the five lineages could be differentiated on the basis of the hybridisation signals observed, however, isolates belonging to lineage IV and III (which have mutations at *oxy*R-37 and *rpoB*-2646 respectively) hybridised to both the *mt* and *wt* probes accordingly. Isolates belonging to lineage I, II and *M. bovis* all have *wt* sequences at these positions and only hybridised to the *wt* probe. Although mutant strains of *oxy*R-37 and *rpoB*-2646 hybridised to both *mt* and *wt* probes differentiation could still be achieved. However, the aim of the macroarray was for each strain tested to hybridise to either the *wt* or *mt* target. To improve the specificity a mismatch was added to the *wt* probe of these two targets in order to destabilise the DNA duplex. Any partial hybridisation of a *mt* strain should be unstable (as the *wt* probe would contain two mismatches) and easily removed during washing. A G-T mismatch was used as this base combination has the weakest bond therefore a *wt* strain should still hybridise to the *wt* probe. This approach was successful with hybridisation occurring at either the *wt* or *mt* probe at each target (Figure 6.3).

Figure 6.3 Hybridisation patterns corresponding to the sSNPs defining each lineage



Orientation of probes from left to right and top to bottom: ink spot; colour development control; 4 blanks; *oxyR*-37 wild type (*wt*); *oxyR*-37 mutant type (*mt*); *katG*-87*wt*; *katG*-87*mt*; 4 blanks; *rpoB*-2646*wt*; *rpoB*-2646*mt*; *rpoB*-3243*wt* and *rpoB*-3243*mt*. NB. Blank spaces were included to extend the length of the macroarray for ease of handling. (A table showing orientation of probes can be found in Material and Methods 2.8.3).

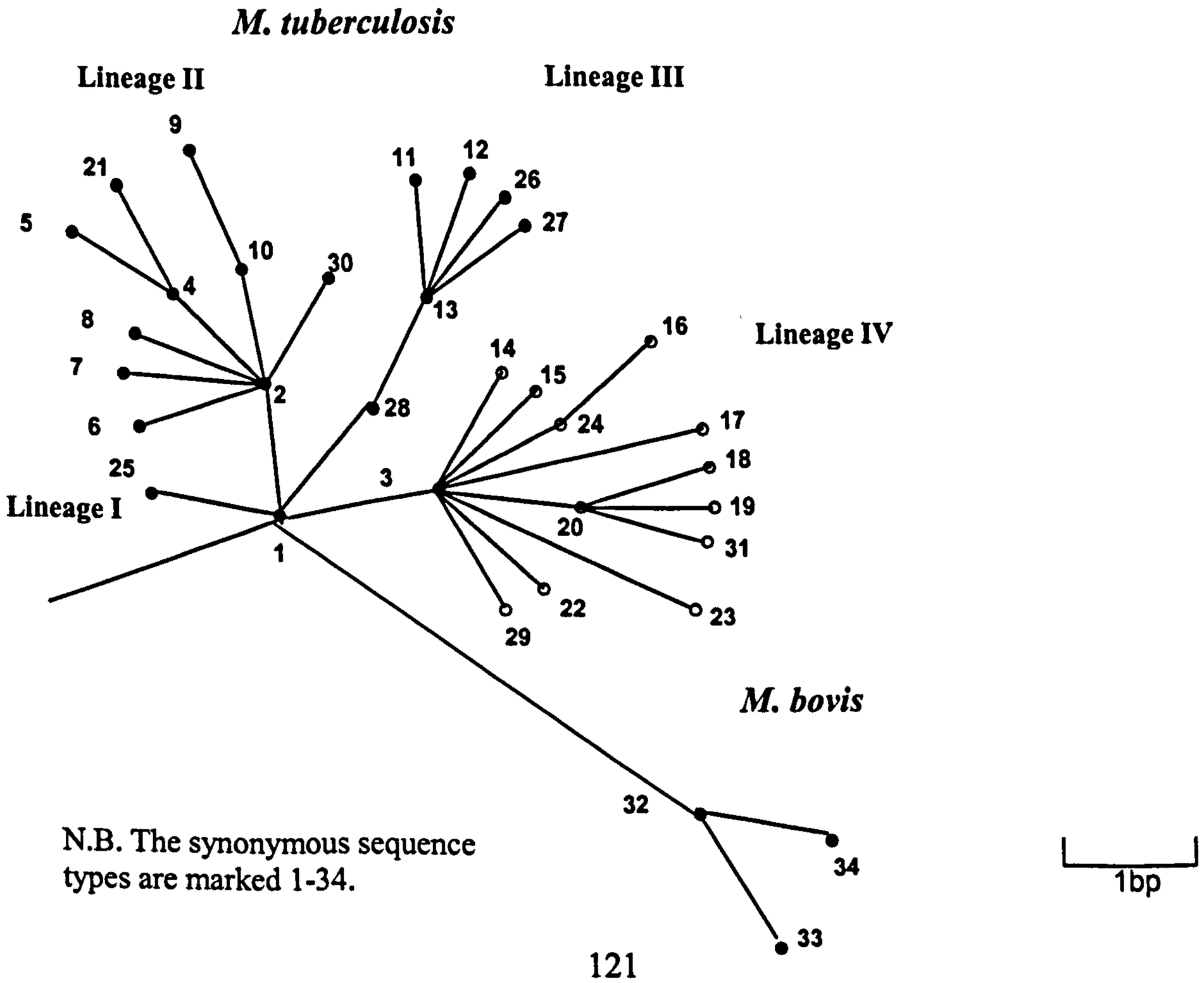
Isolates from each of the five lineages were amplified to demonstrate the hybridisation pattern for each lineage. *M.tb* H37Rv is a lineage II isolate, hence hybridisation at all four *wt* probes. The expected hybridisation patterns are shown in table 6.1.

6.3 Results

6.3.1 VNTR-MIRU analysis of panel 1

The panel of 312 *M.tb* isolates (panel 1) which had been characterised by sSNP analysis into four lineages plus *M. bovis* (Baker *et al.* 2004) was VNTR-MIRU-12 typed. Figure 6.4 displays the five main lineages and sub-lineages based on 225 isolates with complete sequencing data. The remaining 87 isolates could be assigned to one of the 5 main lineages, but not a sub-lineage, using partial sequencing data. VNTR-MIRU profiles were obtained from 309/312 (99.0%) of *M.tb* isolates and 4/4 (100%) of *M. bovis* isolates. These isolates were examined to identify if any single VNTR-MIRU polymorphism or combination of polymorphisms were common to each lineage and could potentially be used to define each lineage.

Figure 6.4 Phylogenetic tree based on sSNPs in 225 isolates with completed sequencing data (Personal communication Dr. L. Baker).



N.B. The synonymous sequence types are marked 1-34.

Analysis of VNTR-MIRU profiles showed that the allelic diversity at certain loci was conserved while others varied considerably (Table 6.2). This was observed both within a lineage and between each lineage.

Table 6.2 Range of repeat numbers observed at loci within each lineage

Locus	Range of repeat no	No of repeats observed within each lineage			
		I	II	III	IV
2	1-5	2	1-4	1-4	1-5
4	1-9	2	1-8	1-2	2-9
10	2-7	3-6	2-7	2-7	2-7
16	0-5	1-4	1-5	2-5	0-4
20	1-3	1-2	1-3	1-2	1-2
23	1-11	5-6	1-6	3-5	2-11
24	1-2	1	1-2	1	2
26	1-12	5-8	1-9	1-12	1-2
27	1-7	1-7	1-4	1-3	1-6
31	1-7	4-6	1-5	4-6	3-7
39	1-4	2-4	1-3	1-4	1-4
40	0-7	1-4	1-7	1-6	0-5
A	1-11	3-4	1-11	1-9	1-10
B	1-10	2	1-3	1-5	1-10
C	1-5	4	1-5	2	1-5

No single polymorphism was able to define each lineage; instead a combination of polymorphisms were used to classify each phylogenetic branch (Table 6.3). For example, in lineage I the majority of isolates (90%) contained three copies of MIRU-10, five or seven copies of MIRU-26 (90%) and three copies of MIRU-39 (90%). None of these markers could be used to define lineage I alone as these repeat numbers were observed in other lineages. By combining these polymorphisms a VNTR-MIRU phylogenetic code was created which could define 80% of lineage I isolates. The same analysis was carried out on the other three lineages and *M. bovis* isolates.

Lineage IV was highly conserved and 92% of isolates contained the VNTR-MIRU phylogenetic code defining their group i.e. two copies of both MIRU-24 and -26. The high allelic diversity in lineage II and III resulted in a low number of isolates containing the lineage-defining polymorphisms, 70% and 52% respectively. *M.tb* H37Rv, which belongs to lineage II, also displayed a lineage II code. *M. bovis* isolates were highly conserved. All isolates contained the three polymorphisms defining this group, however, only 4 isolates were examined therefore analysis of a larger panel of isolates would be needed to validate this VNTR-MIRU code (see section 6.3.2).

Interestingly, the polymorphic codes were almost exclusive. The codes defining lineage II and IV were not observed in any isolates from other lineages. The code for lineage I was seen in one isolate in lineage III and the code for lineage

Table 6.3 VNTR-MIRU loci and repeats common to each of the 4 *M.tb* and *M. bovis* lineages based on VNTR-MIRU-12

Lineage (Total No)	VNTR-MIRU locus[a]	No of repeats observed[b]	No of isolates containing polymorphisms[c]	Total no of isolates with all polymorphisms[d]
Lineage I (20)	10	3	18/20 (90%)	16/20 (80%)
	26	5,7	18/20 (90%)	
	39	3	19/20 (95%)	
Lineage II (168)	26	1,4,5	146/166 (88%)	115/165 (70%)
	31	3	145/163 (89%)	
	39	2	162/165 (98.1%)	
Lineage III (62)	16	4	48/62 (77%)	32/61 (52%)
	39	3	54/62 (87%)	
	40	3	45/61 (74%)	
Lineage IV (62)	24	2	59/61 (97%)	57/62 (92%)
	26	2	58/62 (94%)	
M. bovis (4)	4	3	4/4 (100%)	4/4 (100%)
	10	2	4/4 (100%)	
	40	2	4/4 (100%)	

[a] VNTR-MIRU loci containing little or no variation in each lineage. [b] Most common repeat number observed at these loci and [c] the percentage of isolates containing the repeat number.

[d] The percentage of isolates containing the combination of polymorphisms (described in [a], [b] and [c]) used to classify each phylogenetic branch.

III was seen in one isolate from lineage I and in three isolates in lineage IV. The *M. bovis* code was unique to *M. bovis* and was not observed in any *M.tb* isolates.

6.3.2 Addition of exact tandem repeat (ETR) loci

To try and improve the VNTR-MIRU phylogenetic codes, three additional repeat loci were analysed; exact tandem repeat (ETR) A, B and C from the original 5 loci VNTR method (Frothingham and Meeker-O'Connell 1998) giving 15 loci (VNTR-MIRU-15). The two remaining ETRs D and E were already included in the initial panel of 12-MIRU (ETR-D is MIRU 4 and ETR-E is MIRU 31). Improving the phylogenetic code relies on loci containing little variation within a lineage but enough variation between each lineage to separate them from each other. The main aim of this investigation was to produce a code that was able to define the majority of isolates in each lineage with the least number of markers.

The addition of the three ETRs increased the proportion of definable isolates in all lineages except lineage IV. In lineage I the majority of isolates (95%) contained four copies of ETR-A and all isolates contained 4 copies of ETR-C (Table 6.4). Unfortunately these two markers alone could not be used to define lineage I as this combination was observed in isolates from other lineage, therefore these markers were combined with the original code to determine the best combination of polymorphisms which would define the most isolates. Supplemented with MIRU-39 from the original VNTR-MIRU code, 90% of the isolates in lineage I contained all three polymorphisms. This was a 10%

improvement on the original coding. The same principle was performed for the other lineages.

In lineage II the ETR loci were highly polymorphic. Identifying a single defining allele was difficult; instead a combination of 1 or 2 copies of ETR-B was used and this was able to define 98.1% of the lineage. When this marker was added to the original code only the addition of MIRU-39 was able to improve the percentage of definable isolates; however, these 2 markers were observed in several isolates from other lineages indicating that another marker was required. The addition of MIRU-26 and -31 actually reduced the number of isolates the code could define therefore a new marker was sought. MIRU-16 showed the most potential as 98.2% of isolates contained one, two or three copies of this MIRU. When combined with ETR-B and MIRU-39, 92.7% of isolates were definable which was a 22.7% increase on the original coding.

Two copies of ETR-C were observed in 100% of isolates in lineage III, however, none of the original markers when added to ETR-C improved the phylogenetic code. Instead, five copies of MIRU-23 combined with one copy of ETR-C were able to define 98.4% of isolates. This was a remarkable improvement on the original code which only defined 52% of isolates.

In lineage IV all of the ETRs were very polymorphic and so of little additional value; the original code was therefore retained.

Table 6.4 VNTR-MIRU loci and repeats common to each of the 4 *M.tb* and *M. bovis* lineages based on VNTR-MIRU-15

Lineage	VNTR-MIRU loci (n of isolates with correct repeat no)	Observed in other lineages?
I (20)	39 –3 (19/20 =95%) A – 4 (19/20 =95%) C – 4 (20/20 = 100%) 39-3 + A – 4 + C – 4 = (18/20 = 90%)[a]	2 isolates in lineage IV.
II (168)	16 – 1,2,3 (164/167 = 98.2%) 39 – 2 (160/167 = 95.8%) B – 1,2 (162/165 = 98.1) 16 – 1,2,3 + 39 – 2 + B – 1,2 = (153/166 = 92.7%)	2 isolates in lineage IV
III (62)	23 – 5 (61/62 = 98.4%) C – 2 (62/62 = 100%) 23 – 5 + C – 2 = (61/62 = 98.4%)	1 isolate in lineage IV and 1 isolate in lineage II.
IV (62)	24 – 2 (61/61 = 100%) 26 – 2 (60/62 = 98.8%) 24 – 2 + 26 – 2 = (59/61 = 96.7%)	Not seen in any other lineage
M. bovis (4)	10-2 (4/4 = 100%) 40-2 (4/4 = 100%) C – 5 (4/4 = 100%) 10-2 + 40-2 + C – 5 = (4/4 = 100%)	Not seen in any other lineage

[a] Combined polymorphism can be used to define each lineage

In *M. bovis* only ETR-C was conserved with 100% of isolates containing five copies which was the same level of discrimination as that given with MIRU-4. ETR-C is used to define other lineages whereas MIRU-4 is not therefore it would be beneficial to use ETR-C rather than MIRU-4 in order to keep the total panel of lineage-defining VNTR-MIRU as small as possible to ensure costs and analysis time are kept to a minimum. However this code needs to be applied to the additional *M. bovis* isolates (panel 2) to determine if it is robust as only 4 isolates were tested.

Six isolates possessed VNTR-MIRU polymorphisms associated with more than one lineage. In all cases the isolates possessed the characterising VNTR-MIRU phylogenetic code consistent with the sSNP defined lineage plus polymorphisms associated with another lineage (Table 6.4).

6.3.3 Macroarray analysis

The macroarray (for development see section 6.2.2, for final method see Material and Methods 2.8) was validated using a subset of 46 strains belonging to panel 1. The macroarray and original sSNP results were concordant, i.e. the macroarray correctly identified the lineage of each isolate based on the detection of probe-defined sSNPs.

6.3.4 VNTR-MIRU analysis of panel 2

In order to validate the VNTR-MIRU codes defined in the test population, panel 2 (205 isolates) was VNTR-MIRU-15 typed and analysed for sSNPs with the macroarray. Overall there was 88.9% concordance between the lineage of

isolates defined by the macroarray and VNTR-MIRU. Within the *M.tb* lineages the level of concordance varied from 81.3% to 100% (Table 6.5).

The code for *M. bovis* (2 copies of MIRU-10 and MIRU-40 and 5 copies of ETR-C) was based on only 4 isolates. The analysis was extended to an additional 45 isolates (panel 2) from the human *M. bovis* isolates described in chapter 5 and in Gibson *et al.* 2004. (The original 4 isolates in the Baker *et al.* (2004) panel plus these additional 45 isolates comprised the 49 isolates described in Chapter 5). The applicability of the code was reduced from 100% to 88.9% when the additional 45 isolates were added, indicating that the code was not as robust as it first appeared but still gave a good level of discrimination.

Table 6.5 The number of isolates in panel 2 belonging to each lineage according to the macroarray and its concordance with the VNTR-MIRU phylogenetic codes summarised in table 6.4

Lineage defined by macroarray	n	VNTR-MIRU concordance (%)	Discrepant results (%)
I	16	13/16 (81.3)	3 (18.8)
II	79	68/79 (86.1)	11 (13.9)
III	38	38/38 (100)	0 (0.0)
IV	27	23/27 (85.2)	4 (14.8)
M. bovis	45	40/45 (88.9)	5 (11.1)
Total	205	182 (88.8)	23 (11.2)

6.3.5 Total panel analysed by VNTR-MIRU

Combining panel 1 and 2 gave 518 isolates (465 *M.tb* and 49 *M. bovis*) analysed with both VNTR-MIRU and sSNPs (DNA sequencing or the macroarray). Overall, there was 90.9% concordance between lineage-defined VNTR-MIRU phylogenetic codes and sSNP-defined lineages. The number of definable isolates varied from 86.1 to 99.0% in *M.tb* and was 89.8% in *M. bovis* isolates (Table 6.6).

Table 6.6 The number of isolates belonging to each lineage according to sSNP analysis and its concordance with VNTR-MIRU for all *M.tb* and *M. bovis* isolates tested

Lineage defined by sSNP	n[a]	VNTR-MIRU concordance (%)
I	36	31 (86.1)
II	245	220 (89.8)
III	100	99 (99.0)
IV	88	77 (87.5)
M. bovis	49	44 (89.8)
Total	518	471 (90.9)

[a] Total number of isolates tested (excludes 3 isolates from panel 1 that did not amplify)

6.3.6 Discrepant results

As sSNPs display neutral variation they are unlikely to be under any selection pressures. The sSNPs discussed here are concordant with other genetic and phenotypic groupings (including genetic groups 1 to 3, *M.tb* specific deletion, spoligotyping and IS*6110* families) (Baker *et al.* 2004) and therefore they are likely to be definitive. Any disagreement with the sSNP-defined lineages was considered as discrepant. Overall, 47 (9.1%) discrepant results were observed i.e. the VNTR-MIRU code did not match the lineage defined by sSNP analysis (Table 6.7). Of these, 42 were *M.tb* isolates and the remaining 5 were *M. bovis* isolates. Of the *M.tb* isolates, 27/42 (64.3%) did not match any of the defined VNTR-MIRU codes. These isolates differed in repeat number at one or two loci to one of the VNTR-MIRU phylogenetic codes. In each case this matched the lineage defined by sSNP analysis. For example, five isolates grouped as lineage II discrepancy 'b' had a similar VNTR code to lineage II except that MIRU-39 possessed one copy instead of two copies. Three isolates grouped as lineage II discrepancy 'g' contained five copies of MIRU-16 and three copies of ETR-B as opposed to the predicted one, two or three copies of MIRU-16 and one or two copies of ETR-B. The allelic diversity at the discrepant loci in all lineages displayed a stepwise trend. For example, discrepant isolates 'b' to 'e' in lineage II contained either 1 or 3 copies of MIRU-39 instead of 2 copies and discrepant isolates 'g' and 'i' contained three copies of ETR-B instead of 1 or 2 copies. These discrepancies support the presence of sub-lineages within lineage II.

Twelve *M.tb* isolates had VNTR-MIRU codes for two lineages. Three isolates defined as lineage II by sSNP analysis had the VNTR-MIRU code for both

Table 6.7 Discrepancies observed in each lineage (total population of 518 isolates)

Lineage	Defined MIRU-VNTR loci & repeat no	Discrepancy	Discrepant MIRU-VNTR code(s)[a]	No. of times observed
I	39 –3 A – 4 C – 4	a	**39-2**, **A-3**, C-4 + code for II	1
		b	**39-2**, A-4, C-4 + code for II	2
		c	39-3, **A-3**, C-4	1
		e	**39-4**, A-4, C-4	1
II	16 – 1,2,3, 39 – 2, B – 1,2	a	Codes for lineage II and III present	3
		b	16-3, **39-1**, B-2	5
		c	16-3, **39-1**, B-1	2
		d	16-1, **39-3**, B-1	1
		e	16-3, **39-3**, B-2	4
		f	**16-4**, 39-2, B-2	4
		g	**16-5**, 39-2, **B-3**	3
		h	16-1, 39-2, **B-3**	2
		i	16-3, 39-2, **B-3**	1
III	23 – 5, C – 2	a	**23-3**, C-2	1
IV	24 - 2 26 - 2	a	Codes for lineage I and IV present	6
		b	Codes for lineage II and IV present	2
		c	Codes for lineage III and IV present	1
		d	24-2, **26-1**	2
M. bovis	10-2 40-2 C – 5	a	10-2, 40-2, **C-2**	1
		b	10-2, 40-2, **C-3**	1
		c	10-2, 40-2, **C-4**	1
		d	**10-6**, 40-2, **C-4**	1
		e	10-2, **40-3**, C-5	1

[a] Discrepant loci are highlighted in bold

lineage II and III and nine isolates defined as lineage IV had the code for this lineage plus another *M.tb* lineage code. The phylogenetic code for lineage IV was exclusive and not observed in any other lineage therefore it seems that any discrepant isolate containing the VNTR-MIRU code for lineage IV should be designated a lineage IV strain. However the VNTR-MIRU code for lineage II was not exclusive making it difficult to distinguish which lineage these isolates belong to. Only 3/245 (1.2%) isolates in lineage II had two codes therefore the chances of this occurring are probably low. Interestingly, if the code for lineage II is modified by only counting two copies of ETR-B instead of one and two copies, then the code becomes highly conserved and is not observed in any other lineage. However, the increase in specificity is outweighed by the loss in sensitivity reducing the proportion of isolates defined from 92.7% to 76.3%.

The remaining three *M.tb* discrepant strains were defined as lineage I according to sSNP analysis but all had the lineage II code and a partial code for lineage I. All three had a Beijing spoligotype, i.e. spacers 1-34 absent, 35-43 present, and may represent evolutionary intermediates.

Of the discrepant *M. bovis* isolates all five had different codes. Again a step-wise trend was seen in the discrepant isolates. For example, two, three, four and five copies of ETR-C were observed.

Panel 2 was comprised of 80 *M.tb* isolates from 1998 and 80 from 2004 in order to confirm the phylogenetic analysis was reproducible and not unique to 1998. The number of discrepant isolates in 2004 in each lineage was less in most cases

than the number seen in 1998 (Table 6.8) suggesting that the phylogenetic codes are robust and reproducible through time.

Table 6.8 Number of discrepant isolates in 1998 and 2004

Lineage	No of discrepant isolates in 1998 (n = 389)	No of discrepant isolates in 2004 (n = 80)	Rates Ratio (95% confidence interval)
I	2	3	0.13 (0.02 - 0.81)
II	21	3	1.44 (0.44 – 4.71)
III	1	0	-
IV	9	2	0.93 (0.20 - 4.20)

The difference in the number of discrepant isolates was not significant in any lineage except lineage I. Discrepant isolates were 0.13 times less likely in this lineage in 1998 than in 2004.

6.3.7 Indeterminate isolates

Four isolates not assigned a lineage in the Baker *et al.*, panel because of incomplete MLST data were analysed by VNTR-MIRU to determine whether they could be classified using these codes. Three isolates had the VNTR-MIRU code for lineage II but none of the other codes. The remaining isolate had three copies of MIRU-39 and four copies of ETR-C but five copies of ETR-A rather then four which is seen in the code for lineage 1. This isolate did not have a VNTR-MIRU profile matching any of the other phylogenetic codes. sSNP macroarray analysis confirmed that three of the isolates were indeed from lineage II and the remaining isolate was from lineage I.

6.3.8 Clustering of *M.tb* for molecular epidemiological analysis

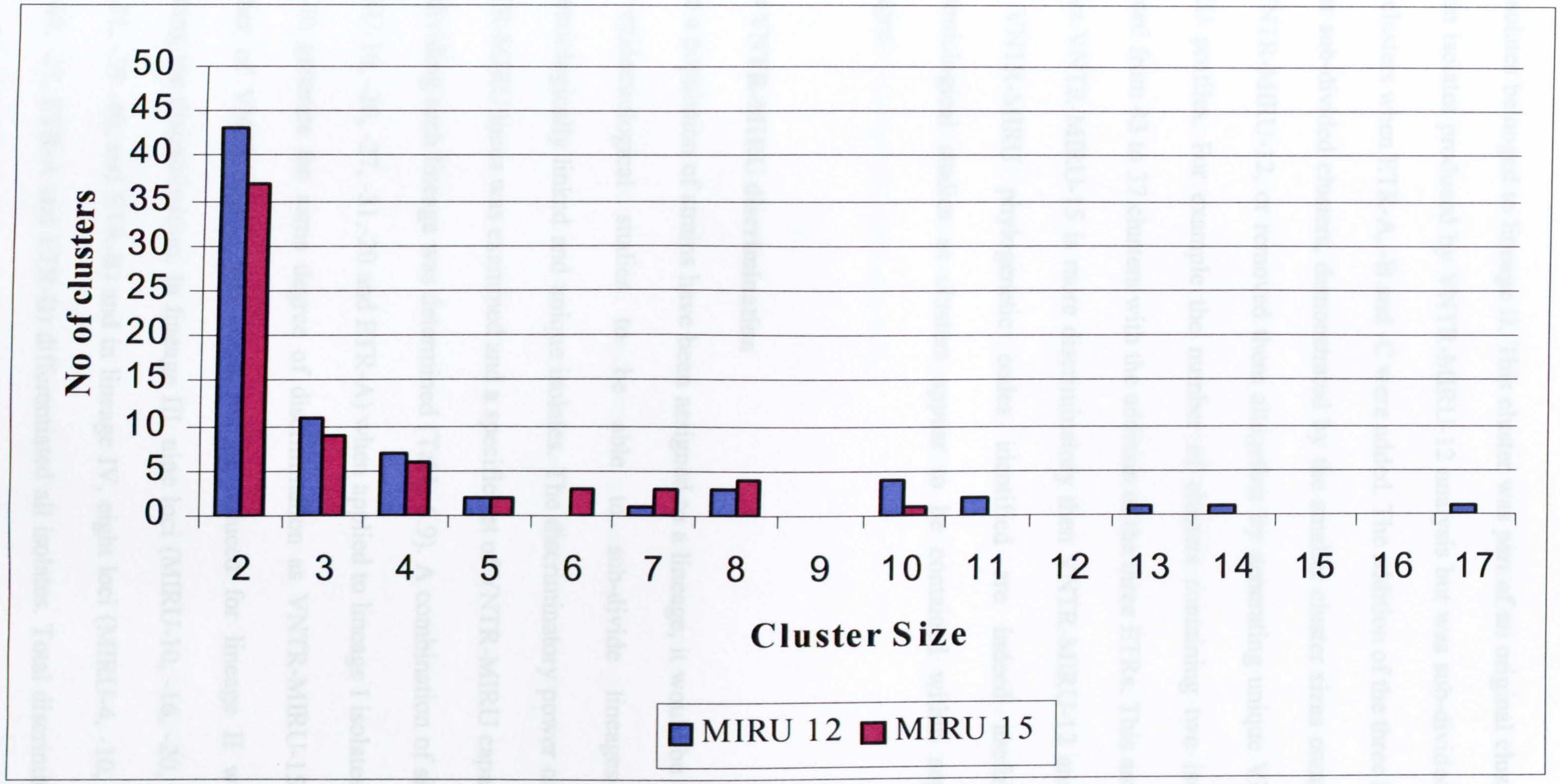
6.3.8.1 VNTR-MIRU-12

VNTR-MIRU profiles from the 518 *M.tb* and *M. bovis* isolates (panel 1 and 2) were compared using a categorical similarity coefficient and displayed using an UPGMA dendrogram. The reason for this analysis was to determine if the VNTR-MIRU phylogenetic codes would be useful for molecular epidemiological investigations i.e. if clusters were confined to one lineage. Cluster analysis based on 12-loci produced 76 clusters containing between 2 and 17 isolates (Figure 6.5). The majority of clusters (56.6%) contained just 2 isolates. Each cluster contained isolates from only one lineage, with 3 exceptions: one cluster of ten isolates contained one isolate from lineage III and the remaining nine isolates from lineage I. One cluster of eleven isolates contained one isolate from lineage IV and the remaining ten isolates from lineage II and finally one cluster of seventeen isolates contained one isolate from lineage I and the remaining sixteen isolates from lineage III.

6.3.8.1 VNTR-MIRU-15

In comparison to the discriminatory power of VNTR-MIRU-12, VNTR-MIRU-15 reduced the number of clusters from 76 to 65. Clusters contained between 2 and 10 isolates and all but one belonged to the same lineage, therefore VNTR-MIRU-15 appears to be more discriminatory than VNTR-MIRU-12. This cluster contained seven isolates, one isolate belonged to lineage IV and the remaining

Figure 6.5 Frequency of cluster size



six isolates belonged to lineage II. This cluster was part of an original cluster of eleven isolates produced by VNTR-MIRU-12 analysis but was sub-divided into two clusters when ETR-A, -B and -C were added. The addition of the three ETRs either sub-divided clusters, demonstrated by the smaller cluster sizes compared to VNTR-MIRU-12, or removed them altogether by generating unique VNTR-MIRU profiles. For example the number of clusters containing two isolates reduced from 43 to 37 clusters with the addition of the three ETRs. This analysis shows VNTR-MIRU-15 is more discriminatory then VNTR-MIRU-12 and that the VNTR-MIRU phylogenetic codes identified are indeed useful for epidemiological studies as clusters appear to be contained within separate lineages.

6.3.9 VNTR-MIRU discrimination

Once a population of strains have been assigned to a lineage, it would be useful for epidemiological studies to be able to sub-divide lineages into epidemiologically linked and unique isolates. The discriminatory power of each VNTR-MIRU locus was examined and a specific set of VNTR-MIRU capable of sub-dividing each lineage was determined (Table 6.9). A combination of six loci (MIRU-16, -26, -27, -31,-20 and ETR-A) when applied to lineage I isolates were able to generate the same degree of discrimination as VNTR-MIRU-15. The number of VNTR-MIRU loci could not be reduced for lineage II without reducing the discrimination. In lineage III, nine loci (MIRU-10, -16, -20, -26, -27, -31, -39 -40, and ETR-B) and in lineage IV, eight loci (MIRU-4, -10, -31, -39, -40, -23, ETR-A and ETR-B) differentiated all isolates. Total discrimination of the *M. bovis* lineage was achieved with eight loci (MIRU-23, -24, -26, -27,

ETR-A, ETR-B and a combination of two VNTR-MIRU drawn from the following; MIRU-4, -31, or –39).

However, if a lower level of discrimination is acceptable then an even smaller number of VNTR-MIRU could be applied which may be of value in resource constrained laboratories. For example by allowing 4 or less discrepancies per lineage (i.e. those isolates that are clustered when using a smaller panel of VNTR-MIRU compared to clustering seen with all 15 VNTR-MIRU); 4, 12, 7, 7 and 6 VNTR-MIRUs would be sufficient to produce the same level of intra-lineage discrimination as the whole panel when applied to lineage I, II, III, IV and *M. bovis* respectively (Table 6.9).

6.3.10 *Mycobacterium africanum*

In Chapter 5 one isolate believed to be an *M. bovis* isolate due to its phenotypic biochemical characteristics but displaying an unusual spoligotypes pattern i.e. contained spacers 40 to 43, was found to be an *M. africanum* strain after deletion typing. This isolate was analysed with the macroarray and found to have a lineage I hybridisation pattern.

Table 6.9 VNTR-MIRU loci used to sub-divide each lineage into true clustered and true unique isolates

Lineage (n)	No. of clusters (range)	No. of clustered isolates	No. of unique isolates	VNTR-MIRU (discrepancies =/<4)[a]	Additional VNTR-MIRU for total discrimination (discrepancies)[b]
I (36)	6 (2- 9)	27	9	16, 26, 27, 31 (4)	+A (1) +A + 20 (0)
II (245)	30 (2-13)	120	125	4, 10, 16, 20, 23, 26, 27, 31, 39, 40, A, C (4)	+B (2) +B+2 (1) +B+2+24(0)
III (100)	15 (2-13)	63	37	10, 16, 26, 31, 39, 40, B (2)	+20 (1) +20+27 (0)
IV (88)	7(2- 4)	19	69	4, 10, 31, 39, A, B, 40 (2)	+23 (0)
M. bovis (49)	5 (2 only)	10	39	23, 24, 26, 27, A, B (3)	A combination of two MIRU from; 4, 31 or 39(0)

[a] Isolates could initially be screened with a smaller panel of VNTR-MIRU in a resource constrained laboratory wishing to produce a reasonable level of intra-lineage discrimination for epidemiological purposes at lower cost. This column shows the maximum number of VNTR-MIRU loci required to produce 4 or less discrepancies i.e. isolates which are clustered when using a smaller panel of VNTR-MIRUs compared to clustering with VNTR-MIRU-15

[b] List of additional VNTR-MIRU loci that would need to be added to the smaller panel to give the same level of intra-lineage discrimination of clustered isolates as VNTR-MIRU-15 typing.

6.4 Discussion

The identification of key phylogenetic markers within populations of isolates is important in order to aid our understanding of *M.tb* complex evolution and pathogenesis. A series of such markers in *M.tb* and *M. bovis* have been previously described by Baker *et al.*, (2004) which defined 5 distinct lineages. VNTR-MIRU is a rapid, high throughput-typing tool which provides useful epidemiological markers for use in TB outbreaks and control programs. If the numeric codes by which VNTR-MIRU types are expressed can also define phylogenetic lineage, very large populations of *M.tb* complex isolates can be studied from an evolutionary perspective as this typing system is beginning to be used routinely in some centres both in the UK and globally.

In this study VNTR-MIRU was successfully applied to a test population to determine phylogenetic codes capable of defining each of the lineages described previously. These VNTR-MIRU codes were then applied to an additional panel using a newly developed portable sSNPs-based macroarray. Developing a macroarray to identify known sSNPs in a population allowed the rapid speciation and identification of key lineages incorporating important TB strain families including, Beijing, CAS/Delhi and EAI and eliminating the need for time consuming more-costly sequencing. The macroarray also proved to be useful in defining the lineage of isolates that cannot be confirmed by sequencing or VNTR-MIRU as shown by the indeterminate isolates which were resolved by the macroarray. The macroarray therefore, would be a useful preliminary screening tool for *M.tb* and *M. bovis* populations.

Overall, the VNTR-MIRU phylogenetic codes were able to define 90.9% of sSNP-characterised *M.tb* and *M. bovis* isolates using a combination of ten loci of the 15 loci examined.

Lineage I accommodates the Beijing family strains. In this study, 86.1% of the Beijing isolates were defined using the VNTR-MIRU lineage I code. It should be noted that even family groups such as Beijing which have been shown to be genetically homogenous can exhibit evolutionary divergence. Although the VNTR-MIRU codes were not as definitive as spoligotyping, *dnaA-dnaN* (Filliol *et al.* 2002, Drobniewski *et al.* 2002, Kurepina *et al.* 1998) and sSNP analysis, the data shown here demonstrates the potential of VNTR-MIRU as a good screening marker for Beijing-type strains. Ferdinand et al., (2004) genotyped spoligotype-defined families using the VNTR-MIRU-12 system. They found that all Beijing isolates could be characterised using a maximum of six MIRUs; as their focus was to characterise all isolates the loci and repeat numbers used were not the same for all Beijing isolates. A single code for the identification of Beijing isolates may therefore be more useful before or in parallel to, discriminating between these isolates.

In terms of allelic diversity the Beijing phylogenetic code loci, MIRU-39 and ETR-A and –C, were very stable (0.05, 0.05 and 0.00 respectively according to the HGI). This low diversity makes them useful markers for phylogenetic purposes (as opposed to a molecular epidemiology tool where high diversity is required) to indicate links between isolates. A study of Russian Beijing isolates

found a similar level of allelic diversity in MIRU 39, ETR-A and ETR-C (0.02, 0.05, 0.02) (V. Nikolayvsky, Personal communication/submitted for publication) suggesting that these markers could be used to define Beijing isolates globally. Unfortunately most Beijing strains characterised by VNTR-MIRU for strain differentiation and/or phylogenetic analysis have used VNTR-MIRU-12 only (Ferdinand *et al.* 2004, Mokrousov *et al.* 2004, Kam *et al.* 2005), therefore, comparison of their populations with our VNTR-MIRU code was not possible.

Lineage III contained the CAS/Delhi family of strains (Filliol *et al.* 2002, Bhanu *et al.* 2002). The VNTR-MIRU code for this family, five copies of MIRU-23 and two copies of ETR-C, defined 99% of isolates, making the differentiation of CAS/Delhi strains within a population simple and specific.

Locus MIRU-24 has previously been used to classify *M.tb* strains into two groups for evolutionary analysis; those containing 1 copy and those containing 2 or 3 copies (Ferdinand *et al.*, 2004). Our VNTR-MIRU code is in agreement with Ferdinand *et al.*, (2004) who found 97.8% East African-Indian (EAI) strains contained 2 copies of MIRU-24. The EAI strains in this study were found in lineage IV however, other strains including *M. bovis* contained 2 copies of MIRU-24 therefore this locus could not be used to define lineage IV alone. The addition of MIRU-26 resulted in this code becoming exclusive and so provided a simple code to define this family.

An important observation was that the *M. bovis* phylogenetic code was unique and not observed in any *M.tb* isolates. Therefore this phylogenetic code is an excellent marker for differentiating between these two species.

The step-wise trend seen in discrepant isolates suggests that evolution is bi-directional in that repeat numbers either increase or decrease over time at all the loci utilised in one of the five phylogenetic codes except for MIRU locus 24. Indeed, certain loci show more allelic diversity than others which may be due different selection pressures acting at different loci. The molecular clocks of some loci may be faster than others and some loci may tolerate more polymorphisms than others. The exact function of MIRUs is unknown; different MIRUs may have distinct functions. One suggestion is that they may perform a regulatory function. Their high variability may be a way for the bacteria to adapt to a new environment. Changes in the copy number of tandem repeats may affect gene regulation (Supply *et al.* 1997), and have been shown to affect the expression of genes involved in adaptive responses (Moxon *et al.* 1994). Variation in copy number may also have an affect on virulence as seen in *Haemophilus influenzae* where changes in tetranucleotide repeat number in genes involved in the biosynthesis of cell wall lipopolysaccharide had an effect on virulence (Moxon *et al.* 1994, van Belkum *et al.* 1997). MIRU have been identified in the intergenic region of *senX3-regX3*, a gene deletion of this two-component system had an effect on the virulence of *M.tb* (Parish *et al.* 2003). Strains of *M.tb* complex other than *M. bovis* BCG were found to have varying numbers of a 77bp MIRU plus a 53bp repeat within the *senX3-regX3* intergenic region, whist *M. bovis* BCG strains only contained copies of the 77bp repeat

(Magdalena *et al.* 1998). The virulence of *M. bovis* BCG is attenuated and is generally regarded as non-pathogenic (although BCG strains can sometimes cause disease), therefore, theoretically there could be a link between the presence/absence of repeat units and virulence although there is no direct evidence to support this.

VNTR-MIRU can simultaneously differentiate strains and define phylogeny as demonstrated in this study. A lineage specific panel of VNTR-MIRU was used to sub-divide each lineage into clustered and unique strains, therefore eliminating the need to analyse all VNTR-MIRU-15 for every isolate. Using a smaller panel of VNTR-MIRU is advantageous in terms of reduced costs and analysis time. In molecular epidemiological studies the greatest discrimination would normally be required therefore isolates could be screened initially with a smaller VNTR-MIRU panel followed by the analysis of clustered isolates with additional VNTR-MIRU for total discrimination.

VNTR-MIRU can characterise a population of *M.tb* and *M. bovis* isolates for phylogenetic/evolutionary purposes. It is also suggested that VNTR-MIRU is able to provide information on sub-lineages and molecular clocks. VNTR-MIRU analysis is a high-throughput, reproducible method, which produces digital results that are readily portable between laboratories and so is suited to phylogenetic studies as well as having a high discriminatory value for molecular epidemiological investigations. Conversely, although genotyping by indexing genetically neutral variation provides a robust phylogeny for *M.tb* and identification of key strain families its value is limited for definitive

epidemiological typing. This study has demonstrated that VNTR-MIRU profiling can be used for rapid classification of *M.tb* and *M. bovis* isolates, and can be used to subdivide *M.tb* isolates into 4 major phylogenetic lineages. This technique could potentially be extended to identify the other members of the *M.tb* complex; *M. africanum, M. microti* and *M. canetti,* and should be investigated.

CHAPTER 7.0 Extending the *Mycobacterium bovis* phylogenetic tree using SNP analysis

7.1 Introduction

Mutations in the base sequence of DNA can occur spontaneously, through errors in DNA replication or recombination events, and through the damaging effects of physical or chemical agents on DNA. When considering phylogeny, two types of single nucleotide polymorphisms (SNP) occur within genes: silent or synonymous (sSNP) and non-synonymous (nsSNP). Synonymous SNPs cause no change in phenotype and variation is genetically neutral whereas nsSNP cause a change in the amino acid of a protein and so this type of variation can be selected for or against. Comparisons of this type of variation in nucleotide sequences provides robust and portable data which is useful for phylogenetic studies (Baker *et al.* 2004, Gutacker *et al.* 2002).

Many methods exist for the identification of DNA mutations but by far the most comprehensive is DNA sequencing. Other methods, such as single stranded conformation polymorphism (SSCP), tend to be used as screening techniques because they do not characterise the mutation observed. SSCP is a low cost and simple method. The genomic or cDNA region of interest is amplified, denatured and the single strands separated by electrophoresis. Mutations cause a change in the tertiary structure of the single stranded DNA through intramolecular interactions. Folding of the molecule affects the mobility of the strand resulting in a banding pattern that differs to the wild-type strain.

SSCP has been successful utilised for detecting mutations causing or attributing to a variety of diseases or disorders in humans such as; hypercholesterolemia (Humphries *et al.* 1997, Pongrapeeporn *et al.* 2001), cancer (Park *et al.* 2005) and cystic fibrosis (Quint *et al.* 2005). This technique has also been used for the identification of mutations in *M.tb*, particularly in the genes associated with drug resistance (Telenti *et al.* 1993, Pretorius *et al.* 1995, Telenti *et al.* 1997, Fang *et al.* 1999b Cardoso *et al.* 2004). Studies using this method for detecting isoniazid and rifampicin resistance found the sensitivity ranged from 87-96% (Telenti *et al.* 1997, Cardoso *et al.* 2004). However, another study which analysed rifampicin resistant and sensitive isolates found that although the specificity was 100% the sensitivity was much lower at 31.4% (Bobadilla-del-Valle *et al.* 2001).

The aim of this study was to expand our understanding of the phylogeny of *M. bovis* using a combination of SSCP and sequencing to identify potential SNPs, in loci analysed in a previous study, which could be used to extend a previously published phylogenetic tree (Baker *et al.* 2004).

7.2 Methods

All available human isolates of *M. bovis* were collected from the HPA-Mycobacterium Reference Unit between 1986 and 2004, and included the 45 *M. bovis* isolates described in chapters 5 and 6 (excluding the 4 *M. bovis* previously sequenced by Baker *et al.* 2004*). M. bovis* isolates collected between 1987-97 had all been kept as cultures on LJ slopes at room temperature while those isolated between 2001-4 were taken from the frozen archives, which were not available at the time of the study detailed in chapter 5. The DNA was extracted

(Material and Methods 2.2.1) and amplification of three regions analysed by the macroarray (excludes the region surrounding the sSNP at nucleotide position 3243 in *rpoB*) was performed (Materials and Methods 2.8.1). These genes, *oxy*R, *katG* and *rpoB*, were targeted as sSNPs had previously been identified in these genes in *M.tb* and *M. bovis*. SSCP was performed on all isolates (Materials and Methods 2.9). One of the four previously sequenced *M. bovis* isolates (Baker *et al.* 2004) which had no mutations in either *oxy*R or *rpoB* but contained the *M. bovis* specific mutation in $katG^{87}$ was used as a control. Any isolate displaying a shift in banding pattern compared to the control was subjected to DNA sequencing (Materials and Methods 2.10) using the same primers as described above (Materials and Methods 2.8.1).

Each *M. bovis* isolate was also VNTR-MIRU-15 typed (Materials and Methods 2.5.2) to establish if they contained the *M. bovis* specific VNTR-MIRU phylogenetic code and to determine if any mutant isolates comprised VNTR-MIRU sub-lineages.

7.3 Results

7.3.1 SSCP and DNA sequencing results

A total of 137/185 (74.1%) *M. bovis* isolates were successfully amplified and subjected to SSCP. Of these, 129 (94.2%) isolates had complete SSCP data on all three genes (*oxy*R, *katG* and *rpoB*) (Table 7.1). Overall 28 shifts in band position were identified on polyacrylamide gels in 16 isolates: 12 band changes in *oxy*R, 15 in *katG* and one in *rpoB*. An example of a band shift observed in *katG* is shown in figure 7.1.

Table 7.1 Number of isolates with SSCP data and the number of isolates containing potential mutations in each gene.

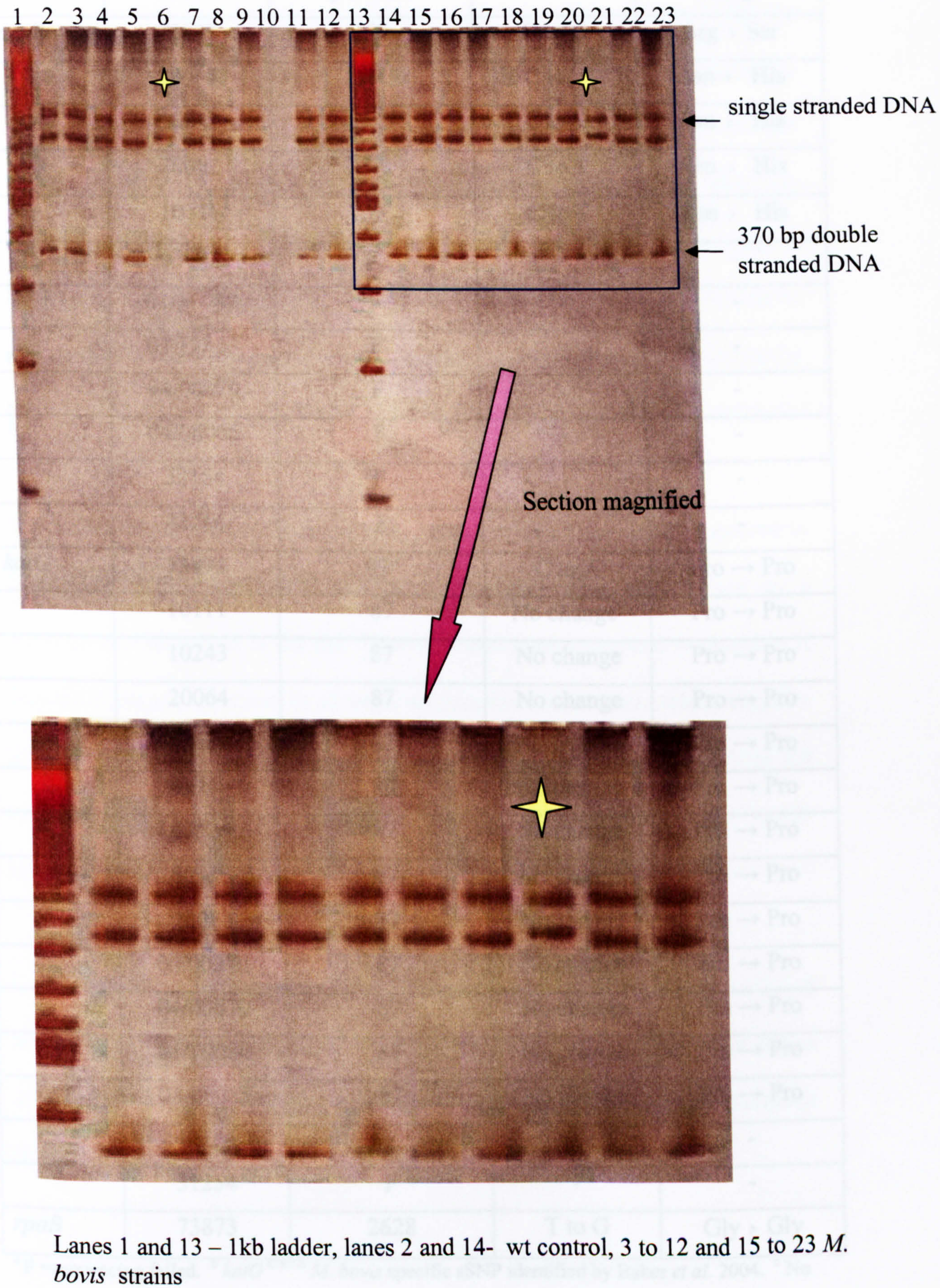
Gene (nt region)	No of isolates with SSCP data	No of isolates with band shifts
*oxy*R (1-311)	135	12
katG (58-428)	132	15
rpoB (2588-2746)	133	1
Total no isolates containing a band shift[a]	129	16

[a] Some isolates contained a band shift in more then one gene region

Of the 12 isolates with band shifts in *oxyR*, seven isolates could not be sequenced and the remaining five displayed one of the following mutations: four had a nsSNP at nucleotide position 37 and one had a nsSNP at position 159 (Table 7.2). In *rpoB*, one isolate had a sSNP at nucleotide 2628. In *katG*, one isolate had the *M. bovis* specific sSNP at *katG*[87] identified by Baker *et al.* (2004), while 12 isolates did not contain this sSNP i.e. were *M.tb*-like, and two isolates could not be sequenced.

Isolates that could not be sequenced were run on a 1.5% agarose gel to check for amplification and that the product was the correct size. In all cases a product of the correct size was observed. The PCR products of these isolates were then purified to remove any substances that may have inhibited the sequencing reaction. Two purification methods were used; standard PEG purification

Figure 7.1 Examples of band shifts on a SSCP gel in k*atG*



Lanes 1 and 13 – 1kb ladder, lanes 2 and 14- wt control, 3 to 12 and 15 to 23 *M. bovis* strains

✦ Lanes containing band shifts

Table 7.2 SNPs identified in each gene observed after sequencing

Gene	Lab n of isolate	Nucleotide position	Base change	Amino Acid change
*oxy*R	73873	159	G to C	Arg › Ser
	20064	37	C to T	Asn › His
	10243	37	C to T	Asn › His
	70082	37	C to T	Asn › His
	10111	37	C to T	Asn › His
	63524	F[a]	-	-
	01/10080	F	-	-
	03/11343	F	-	-
	04/00870	F	-	-
	04/08065	F	-	-
	32154	F	-	-
	42764	F	-	-
katG	73873	87[b]	C to A	Pro → Pro
	10111	87	No change[c]	Pro → Pro
	10243	87	No change	Pro → Pro
	20064	87	No change	Pro → Pro
	42680	87	No change	Pro → Pro
	41963	87	No change	Pro → Pro
	63524	87	No change	Pro → Pro
	60936	87	No change	Pro → Pro
	70082	87	No change	Pro → Pro
	00/00970	87	No change	Pro → Pro
	04/00870	87	No change	Pro → Pro
	01/10080	87	No change	Pro → Pro
	04/08065	87	No change	Pro → Pro
	42764	F	-	-
	31254	F	-	-
rpoB	73873	2628	T to G	Gly › Gly

[a] F =sequencing failed. [b] *katG* $^{C\,87\,A}$ *M. bovis* specific sSNP identified by Baker *et al.* 2004. [c] No change in base compared to *M.tb* H37Rv, i.e. *katG* $^{C\,87\,A}$ *M. bovis* specific sSNP not observed.

(Material and Methods 2.11) and a QIAquick PCR purification kit (Qiagen, West Sussex, UK) which was performed according to the manufacturer's instructions. After purification each sample was sequenced but no result was seen. Next, the PCR products were diluted down 1 in 1000µl to dilute away any possible inhibitors. The PCR reaction was then repeated with the diluted DNA solution. Again no sequencing results were observed suggesting that possibly inhibitors were not the source of the problem. This suggested that the denaturation stage of 20sec may not have been long enough to separate the DNA strands due to increased secondary structure in these samples. To compensate, the denaturation step was increased to 1 min at 96°C, however no sequencing result was obtained.

Ten isolates (two isolates from *oxy*R and eight isolates from *rpoB*) appeared to display slight band aberrations which may possibly have been caused by a mutation or a distortion in the gel. These isolates were sequenced to ascertain if they did contain a mutation. In all cases no mutations were found which suggests the DNA fragments may have been distorted for technical reasons such as the temperature of the gel during electrophoresis becoming too hot or an air bubble in the gel.

Overall four different mutations were identified; two in *oxy*R, one in *katG* and one in *rpoB*. Of these mutations, two were nsSNP i.e. caused a change in amino acid while the remaining two mutations were sSNP i.e. did not cause a change in amino acid.

Interestingly one isolate (lab no 73873) which contained the *katG*^{87} mutation also had a mutation in the *oxyR* and *rpoB* genes. Although a sSNP was found at *katG*^{87}, this mutation could not have accounted for the band shift seen on the SSCP gel as the *M. bovis* control strain also contained this sSNP, therefore another mutation must exist. A mutation may have been present near the primers as the sequencing data obtained did not cover these regions.

The mutation at *oxyR*^{37} (C to T) had previously been seen in *M.tb* isolates (Baker *et al.* 2004) suggesting that either, *M. bovis* can also contain this sSNP or that these isolates were not *M. bovis*. To check that the isolates containing *oxyR*^{37} and those isolates not containing the *katG*^{87} mutation were actually *M. bovis*, spoligotyping and Genotype MTBC typing (HAIN Lifesciences, Nehren, Germany) were performed (Table 7.3). Genotype MTBC typing, which is based on polymorphisms in the *gyr*B gene, was performed according to the manufacturer's instructions. Three of the four isolates containing *oxyR*^{37} had *M.tb* spoligotypes and the remaining isolate had a spoligotype which looked similar to an *M. bovis* spoligotype, but contained spacer 39 which is normally missing. This spoligotyping pattern also had similarities to the other three isolates containing the *oxyR*^{37} sSNP therefore it may not be *M. bovis* but actually an *M.tb* isolate. Genotype MTBC typing identified these isolates were all *M.tb* although this technique cannot differentiate between *M.tb, M. africanum* II and *M. canetti* (personal communication, M. Yates). Interestingly all these isolates contained the VNTR-MIRU code for *M.tb* lineage IV.

Table 7.3 Spoligotype results for isolates with sequencing data

Lab No	SNP identified	Spoligotype (sub-species)	VNTR-MIRU lineage?[a]
10111	*OxyR*[37], no *katG*[87]	■■•■■■■■■■■■■■■■■■■••■■■■■■■••••■•■■■■■■■■■ (*M.tb*)	IV
10243	*OxyR*[37], no *katG*[87]	■■•■■■■■■■■■■■■■■■■••■■■■■■■••••■•■■■■■■■■■ (*M.tb*)	IV
20064	*OxyR*[37], no *katG*[87]	■■•■•■■■■■■■■■■■■■■••■■■■■■■••••■•■■■■■■■■■ (*M.tb*)	IV
70082	*OxyR*[37], no *katG*[87]	■■■■■■■■■■■■■■■■■■■■■■■■■■■■••••■•■■■■■•••• (*M.tb* /*M. bovis*)	IV
73873	*OxyR*[159], *katG*[87], *rpoB*[2628]	■■■••••■■■■■■■■■■■■■■■••••••••••••■■■■■■■■■ (*M.tb*)	None
41963	no *katG*[87]	■■■■■■■■■■■■■■■■■■■■••••••••••••••••••••••• (*M.tb*)	II
42680	no *katG*[87]	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■••••■■■■■■■ (*M.tb*)	II
60936	no *katG*[87]	•■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■••••■■■■■■■ (*M.tb*)	II
63524	no *katG*[87]	■■■•■■■■■■■■■■■■■■■■■•••■••••••■••••■■■•■■■ (*M.tb*)	II
00/00970	no *katG*[87]	•••■■■■•••••••••••••••■■■■■■■■■■■■■■■■••••• (*M. bovis*)	None
01/10080	no *katG*[87]	N/A	None
04/00870	no *katG*[87]	■■■••••••■■■••••••••••••■••■■■■■■■••■■•■■■■ (*M. africanum*)	None
04/08065	no *katG*[87]	■■■■■■■•••••■■■■••••••••■■■■■■■■■■■■•••■■■■ (*M. africanum*)	None

[a] Isolates containing *M.tb* or *M. bovis* VNTR-MIRU phylogenetic codes

Of the 12 isolates that did not contain the *katG*^{87} sSNP, four also had the *oxyR*^{37} sSNP (their spoligotypes and Genotype MTBC results are described above), four isolates had spoligotyping patterns and Genotype MTBC results suggestive of *M.tb* and four isolates were identified as *M. africanum* I by Genotype MTBC typing. Of these latter four, one had an *M. bovis* spoligotype, two had spoligotypes resembling *M. africanum* and one had no spoligotype. Interestingly, the four isolates with *M.tb* spoligotypes all had lineage II VNTR-MIRU profiles. The four isolates labelled as *M. africanum I* by Genotype MTBC typing did not have a *M.tb* or *M. bovis* phylogenetic code. The two remaining SNPs identified, *oxyR* $^{G\ 159\ C}$ and *rpoB*$^{T\ 2628\ G}$, were both found in one isolate (lab no 73873). This isolate had a *M.tb* spoligotype but no Genotype MTBC typing result was available.

Overall the spoligotyping and Genotype MTBC data suggests that 13 out of 16 isolates displaying band shifts were not actually *M. bovis* isolates, the remaining three isolates had no sequencing data for *oxyR* or *katG* and also had no spoligotyping or Genotype MTBC result. These isolates were collected from the *M. bovis* archives at the MRU however on closer inspection their phenotypic and biochemical characteristics were not always typical for *M. bovis*; three isolates were sensitive to pyrazinamide (two of which had an *M.tb* spoligotype and one had no spoligotype) and three isolates were TCH (thiophen-2-carboxylic acid hydrazide) + or +/- (two of which had an *M.tb* spoligotype and one had no spoligotype). The phenotypic and biochemical characteristics of a sample of 20 isolates with no band shift were examined to determine if these isolates all contained typical *M. bovis* features: all 20 isolates had typical *M. bovis*

characteristics i.e. microaerophilic, TCH negative, pyrazinamide resistant and grew better on pyruvate than glycerol Lowenstein-Jensen slopes.

7.3.2 VNTR-MIRU analysis

Each isolate was VNTR-MIRU-15 typed to determine if the VNTR-MIRU phylogenetic code for *M. bovis* identified in chapter 6 was present in all strains. A total of 131/138 (94.9%) isolates had complete VNTR-MIRU data (N.B. one isolate that would not amplify for SSCP analysis was included in the VNTR-MIRU analysis as it was one of the 45 original *M. bovis* isolates hence a total of 138). The *M. bovis* code; 2 copies of MIRU-10, 2 copies of MIRU-40 and 5 copies of ETR-C were found in 97/131 isolates (74.0%). None of the 16 isolates which contained a shift of band position contained the *M. bovis* phylogenetic code which is not surprising as according to the combined data from, SNP analysis, spoligotyping and Genotype MTBC typing, 13 of these isolates did not appear to be *M. bovis*. If these isolates were removed from the analysis 97/118 (82.2%) isolates contained the *M. bovis* phylogenetic code.

7.4 Discussion

Genes with sSNP previously identified in *M.tb* and *M. bovis* as having lineage defining genotypes were targeted as possible 'hot spots' for the identification of further mutations in *M. bovis*. SSCP analysis was successfully utilised as a screening method for the detection of mutations in these isolates which were subsequently characterised by direct DNA sequencing.

Of the 137 isolates analysed, 16 isolates had band shifts compared to the wild-type control using SSCP analysis. From the 13 isolates with complete sequencing data, four different SNPs were found in the three genes examined of which two were non- synonymous and two were synonymous mutations. The two sSNP at *oxyR*^{37} and *katG*^{87} had been identified before in *M.tb* and *M. bovis* respectively (Baker *et al.* 2004). The two nsSNP at oxyR^{159} and rpoB^{2628}, where found in an isolate with typical phenotypic and biochemical characteristics of *M. bovis* and the sSNP at *katG*^{87} suggestive of *M. bovis* but a spoligotype pattern characteristic of *M.tb*. This isolate may have been an *M. bovis* strain but could also represent an intermediate between *M.tb* and *M. bovis*. These nsSNPs could therefore not be added to the Baker *et al.* (2004) tree.

It is also unclear from this data which of the mutations at *oxy*R^{159} and *rpoB*^{2628} came first in the evolutionary pathway as no intermediate containing just one of these mutations were observed. To determine the order of acquisition of these SNPs additional isolates need to be examined.

There are however limitations to adding these non-synonymous SNPs to the phylogenetic tree. Baker *et al.* (2004) constructed their tree using synonymous SNPs which are neutral variation and therefore unlikely be selected for and against, creating an unambiguous tree, whereas the opposite may be true for nsSNP.

In summary, from the 13 isolates with SNP data, eight appeared to be *M.tb* according to the combined data from SNP analysis, spoligotyping, VNTR-MIRU

analysis and Genotype MTBC typing. The remaining five isolates did not have concordant results between the four typing methods and so may represent intermediates within the *M.tb* complex. To ascertain the identity of these isolates further work is required for example by deletion typing which separates the *M.tb* complex based on the presence or absence of regions of difference in the genome (Brosch *et al.* 2002, Mostowy *et al.* 2005).

The sensitivity of the *M. bovis* VNTR-MIRU phylogenetic code decreased with the addition of more *M. bovis* isolates from 88.9% (45 isolates from chapter 6) to 74.0% (total 131 isolates); however the 13 of the 16 isolates with mobility shifts did not appear to be *M. bovis*. Removing these isolates from the analysis increased the number of definable isolates to 82.2%.

A total of nine isolates with band shifts in *oxy*R and *katG* could not be sequenced. Various purification techniques were used but no interpretable sequencing results were obtained. Other reasons that could explain the poor sequencing results include, for example, the fact that PCR is exponential whereas sequencing is not. Inefficient primers can still work for PCR whereas they will affect the efficiency of a sequencing reaction. Primers can also anneal to a mismatched priming site. The mutation could have been located in the primer sequence and so affected the binding of the primers in the sequencing reaction. These primers were successfully used for VNTR-MIRU typing, however during PCR if a primer manages to anneal once to a mismatch region then the extension product will then be a perfect match for subsequent cycles. This is not the case in sequencing reactions and so a mismatch is likely to affect the result.

Overall 48 extracted isolates would not amplify using the original macroarray primers. Almost all of these isolates were kept on LJ slopes at room temperature for, in most cases, over a decade. This suggests that although PCR can be performed on non-viable isolates it is likely the quality of DNA will deteriorate over time if the culture is not kept under suitable conditions i.e. frozen at –70°C which is now standard practice.

Many factors can affect the results of SSCP such as the running temperature of the gel, length of the polyacrylamide gel, running time and speed, and length of the PCR fragments (Humphries *et al.* 1997). In this study a magnetic stir was used during electrophoresis to prevent the occurrence of hot spots in the gel. The gel was also run overnight at a low voltage to minimise overheating. Large gels (20cm) were used, maximising running time to allow adequate band separation. An optimal length of fragments analysed by SSCP of between 150 and 200 nucleotides has been suggested (Humphries *et al.* 1997). In this study only one of the regions analysed was between these boundaries (*rpoB* length = 159bp) therefore mutations may have been missed in the other two genes affecting the sensitivity of the technique. It is believed that changes in conformation will not be picked up as easily the longer the fragment is. The sensitivity of SSCP is this study was not known as only isolates with band shifts were sequenced. The specificity of the technique was 100% although it is worth noting that the mutation detected in one isolate with a band shift could not be found upon sequencing; however this isolate had incomplete sequencing data. This suggests that this isolate was not a false positive result but rather the mutation was in a

region with no sequencing data. Further investigation with additional primers would be required to confirm this.

This study found that SSCP is a cost effective, rapid, method for screening *M. bovis* isolates for mutations. The gene regions in this study were targeted because mutations in *M.tb* and *M. bovis* had previously been identified in these areas (Baker *et al.* 2004). Although the SNPs identified appeared not to be in *M. bovis* isolates the approach was successful as all genes analysed contained mutations. Therefore the other genes analysed by Baker *et al.* (2004) should be examined for additional mutations which may further expand the *M. bovis* phylogenetic tree.

CHAPTER 8.0 Conclusions and future work

8.1 Transmission of *M.tb* and *M. bovis*

Molecular typing is a useful tool for the identification of related cases that may not be linked through traditional epidemiological or contact tracing methods. Identical fingerprints, when using methods with a high degree of discrimination such as RFLP-IS*6110*, are thought to represent recently infected cases and can be targeted for epidemiological investigation. Unique fingerprints are not thought to be epidemiologically related and may have arisen either from the reactivation of a latent infection or the immigration of patients who have acquired infection from abroad (Alland *et al.* 1994, Maguire *et al.* 2002).

The data obtained from the molecular epidemiological study of TB in England found low levels of clustering (21%), this was comparable to a study of TB in London between 1995-1997 which found 23% of cases were clustered (Maguire *et al.* 2002). The low number of potentially linked cases suggests low levels of recent transmission in England. The majority of TB cases may reflect reactivation and importation of disease through recent immigrants from high incidence countries. Risk factors for clustering were identified and included cases with previous treatment and pulmonary smear negative disease. A case control study demonstrated that homelessness associated with alcohol dependence was also a risk factor for clustering.

High levels of clustering (73%) were observed in Harare, Zimbabwe suggesting high rates of recent transmission, however, some caution is needed in this

interpretation as RFLP-IS*6110* fingerprinting could not be performed on these strains. It is likely that a degree of false clustering was present as the discriminatory value of spoligotyping and VNTR-ETR is lower then that of RFLP-IS*6110*. Surprisingly, no risk factors for clustering were observed even with a substantially high level of HIV co-infection (74%). Again this may have been due to the lower discriminatory power of spoligotyping and VNTR-ETR, merging epidemiologically linked and unlinked cases and affecting the identification of risk factors. The two transmission studies of England and Harare could not be compared directly as slightly different molecular techniques were used. However, a high rate of clustering was observed in Harare where TB and HIV are endemic, whereas relatively low rates of clustering were observed in England where the incidence of both TB and HIV infection are low. HIV infection increases the susceptibility to exogenous TB infection and the rapid onset of disease after infection. High levels of HIV in a population could lead to a higher prevalence of infectious TB cases over a shorter period of time therefore this may have contributed to the high levels of clustering observed in Harare.

A comparison of spoligotyping families observed in the two contrasting settings, London (a low incidence population) and Harare (a high incidence population) found a greater diversity of spoligotype strain families in London. Just two main families, the T and the LAM family were observed in Harare, whereas in London 13 large families were present (those containing 20 or more isolates); a probable reflection on the greater international diversity of the population in London.

Of all the strain types present in these two contrasting settings it was found that the T1 strain was ubiquitous. This is not surprising as the T family is thought to

be a relatively old genotype (Filliol *et al.* 2002) and the T1 strain is found worldwide (Filliol *et al.* 2003). *M. Africanum* strains are commonly associated with TB patients in Africa (Niemann *et al.* 2002) however no spoligotype patterns from this species were observed in Zimbabwean isolates. Conversely 7 *M. Africanum* strains were observed in London isolates although this species has previously been seen in South-East of England (Grange and Yates 1989). To conclude spoligotype signatures can be used to screen populations allowing the identification of clades/strain families which may be useful for both public health action and monitoring the global spread of TB.

A comparison of typing techniques using *M.tb* isolates from England between 1st January 1998 and 31st March 1999 demonstrated that VNTR-MIRU analysis displayed high levels of discrimination. However, in a HCN population (isolates with five or more bands) the gold standard RFLP-IS*6110* still provided the greatest level of discrimination. A combination of RFLP-IS*6110* and VNTR-MIRU were more discriminatory than the commonly used combination of RFLP-IS*6110* and spoligotyping suggesting the former are most useful for public health and contact tracing purposes. Further investigation is required to determine if VNTR-MIRU can replace RFLP-IS*6110*, as this would allow a prospective typing service which currently is not possible due to the slow turnover time of RFLP. A total of 79 variable regions have been identified within the *M.tb* H37Rv genome (http://minisatellites.U.psud.fr) many of which have not been assessed for their variability in repeat number between strains. Furthermore other variable regions may exist in clinical strains of *M.tb* which may provide the additional level of discrimination required.

Unfortunately RFLP-IS*6110* and VNTR-MIRU could not be performed on the Zimbabwean isolates due to contaminated cultures and the low yields of DNA obtained. However from the evidence presented in the comparative study of typing techniques it is likely that the level of clustering would have been reduced with the application of RFLP-IS*6110* and VNTR-MIRU. Relatively few molecular epidemiological studies using VNTR-MIRU have been carried out in high incidence areas particularly in Africa. As VNTR-MIRU is a PCR-based technique it overcomes the problems of requiring viable cultures, therefore this should be addressed in future to determine how discriminatory VNTR-MIRU typing is in areas where TB and HIV are endemic.

The majority of *M. bovis* infections in humans in the UK were most likely due to reactivation of a latent infection acquired pre-milk pasteurisation rather than primary infection as almost three-quarters of cases were over the age of 50. Spoligotyping demonstrated the most prevalent strain in humans (Type 9) was also the most prevalent in cattle populations. Thirteen spoligotypes were unique to humans and had not previously been observed in UK cattle. Due to the high levels of reactivation, these isolates may represent strains circulating in UK cattle over 50 years ago.

VNTR-MIRU using 15 loci provided the highest discriminatory power when using a single technique for analysing human *M. bovis* isolates in the UK. Furthermore it also proved to be the best technique for discriminating between *M. bovis* isolates belonging to spoligotype 9. In future VNTR-MIRU-15 should

be considered alone or in combination with spoligotyping for epidemiological studies of human *M. bovis*. It is also possible that this technique will be of value in determining transmission dynamics in *M. bovis* in cattle and other mammals.

8.2 Phylogeny of *M.tb* complex

VNTR-MIRU provided a rapid, robust tool for the phylogenetic analysis of *M.tb* and *M. bovis*. Over 90% of *M.tb* and *M. bovis* isolates were characterised by defined VNTR-MIRU phylogenetic codes, which could also differentiate between *M.tb* and *M. bovis* strains. VNTR-MIRU was not only useful for phylogenetic studies but could simultaneously strain type isolates and was therefore beneficial for epidemiological purposes. Future work should concentrate on improving the phylogenetic codes by analysing additional variable regions both to improve existing codes. This is particularly important in Beijing strains (lineage I) where the phylogenetic code defined the least number of isolates compared to the other *M.tb* lineages. Furthermore, improving on these codes will enable further study of sub-lineages within the four lineages of *M.tb* and *M. bovis*. This technique should also be extended to the other members of the *M.tb* complex including; *M. africanum*, *M. microti* and *M. canetti* to determine whether a phylogenetic code for these species could be identified.

The step-wise trend seen in discrepant isolates suggests that the evolution of VNTR-MIRU may be bi-directional in that the number of repeats at each locus can increase or decrease over time. This should be investigated further by analysing serial isolates from patients. Any changes identified may determine if repeat number can increase, decrease or change in both directions over time.

Little variation in VNTR-MIRUs was observed in serial isolates from patients in South Africa (Savine *et al.* 2002) i.e. only one change at one locus was observed in 56 patients in up to a six year period. However, this is a relatively small sample and so further analysis of serial isolates from both high and low incidence settings should be analysed to elucidate this theory.

Single stranded conformation polymorphism (SSCP) was successfully used as a screening technique to identify mutations. Band shifts were identified in 16 isolates recorded as *M. bovis*. Overall, four mutations were identified by direct sequencing, of which two, *oxy*R[37] and *katG*[87], had been previously identified.

One isolate contained the remaining two mutations at oxyR[159] and rpoB[2628] however it is unclear from this data which of the mutations appeared first in the evolutionary pathway and indeed how these SNPs would fit into the Baker *et al.* (2004) phylogenetic tree as the identity of this isolate was unclear. Typing data suggested it could be an intermediate between *M. bovis* and *M.tb*.

Indeed of the 13 isolates with band shifts and complete sequencing data, none appeared to be *M. bovis* isolates. The identities of eight isolates were concordant by all typing methods (SNP analysis, spoligotyping, VNTR-MIRU analysis and Genotype MTBC typing) and appeared to be *M.tb* isolates. The remaining four isolates were not concordant between these typing techniques and could represent intermediates within the *M.tb* complex. Future work would therefore involve defining these isolates. Deletion typing has been used to differentiate between members of the *M.tb* complex (Brosch *et al.* 2002). This technique is

based on presence or absence of regions of difference in the genome. Recently additional deleted regions have been identified in *M. bovis* and *M. bovis*-like isolates using an Affymetrix GeneChip system (Mostowy *et al.* 2005). This approach has further differentiated the *M. bovis* lineage and could be used to analyse these discrepant isolates to determine if they are indeed sub-species within *M. bovis*.

In future the three genes partially analysed in this study; *oxy*R, *katG*, and *rpoB* should be targeted in full to determine if SNPs can be identified in the isolates confirmed as *M. bovis*. In addition the remaining four genes; *ahpC*, *pncA*, *rpsL* and *gyrA* analysed by Baker *et al.* (2004), which were found to contain sSNPs in *M.tb*, could contain sSNPs for the construction of a *M. bovis* phylogenetic tree and so should be investigated.

REFERENCES

Agerton, T. B., S. E. Valway, R. J. Blinkhorn, K. L. Shilkret, R. Reves, W. W. Schluter, B. Gore, C. J. Pozsik, B. B. Plikaytis, C. Woodley, and I. M. Onorato. 1999. Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis*, across the United States. Clin. Infect. Dis. **29**:85-92.

Alland, D., G. E. Kalkut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom. 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N. Engl. J. Med. **330**:1710-1716.

Aranaz, A., E. Liebana, A. Mateos, L. Dominguez, D. Vidal, M. Domingo, O. Gonzolez, E. F. Rodriguez-Ferri, A. E. Bunschoten, J. D. Van Embden, and D. Cousins. 1996. Spacer oligonucleotide typing of *Mycobacterium bovis* strains from cattle and other animals: a tool for studying epidemiology of tuberculosis. J. Clin. Microbiol. **34**:2734-2740.

Baker, L. V., T. J. Brown, M. C. Maiden, and F. A. Drobniewski. 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg. Infect. Dis. **10**:1568-77.

Behr, M. A., S. A. Warren, H. Salamon, P. C. Hopewell, d. L. Ponce, C. L. Daley, and P. M. Small. 1999. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. Lancet. **353**:444-449.

Bhanu, N., D. van Soolingen, J. van Embden, L. Dar, R. Pandey, and P. Seth. 2002. Predominace of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. Tuberculosis (Edinb.). **82**:105.

Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. T. Crawford, J. M. Musser, and B. N. Kreiswirth. 1996. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA. **275**:452-457.

Bifani PJ, B. Mathema, Kurepina NE, and B. N. Kreiswirth. 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. 45-52.

Bloom, B. R. Tuberculosis: pathogenesis, protection and control. Washington DC: ASM Press, 1994, pp 47-60.

Bobadilla-del-Valle, M., A. Ponce-de-Leon, C. Arenas-Huertero, G. Vargas-Alarcon, M. Kato-Maeda, P. M. Small, P. Couary, G. M. Ruiz-Palacios, and J. Sifuentes-Osornio. 2001. *rpoB* gene mutations in rifampin-resistant *Mycobacterium tuberculosis* identified by polymerase chain reaction single-stranded conformational polymorphism. Emerg. Infect. Dis. **7**:1010-1013.

Braden, C. R., G. L. Templeton, M. D. Cave, S. Valway, I. M. Onorato, K. G. Castro, D. Moers, Z. Yang, W. W. Stead, and J. H. Bates. 1997. Interpretation of

restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. J. Infect. Dis. **175**:1446-1452.

Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. U.S.A. **99**:3684-3689.

Cardoso, R. F., R. C. Cooksey, G. P. Morlock, P. Barco, L. Cecon, F. Forestiero, C. Q. Leite, D. N. Sato, M. L. Shikama, E. M. Mamizuka, R. D. Hirata, and M. H. Hirata. 2004. Screening and characterization of mutations in isoniazid-resistant *Mycobacterium tuberculosis* isolates obtained in Brazil. Antimicrob. Agents. Chemother. **48**:3373-3381.

Cavanagh, R., M. Begon, M. Bennett, T. Ergon, I. M. Graham, P. E. de Haas, C. A. Hart, M. Koedam, K. Kremer, X. Lambin, P. Roholl, and D. D. Soolingen. 2002. *Mycobacterium microti* infection (vole tuberculosis) in wild rodent populations. J. Clin. Microbiol. **40**:3281-3285.

Claij, N. and H. te Riele. 1999. Microsatellite instability in human cancer: a prognostic marker for chemotherapy? Exp. Cell. Res. **246**:1-10.

Clifton-Hadley, R. S., J. W. Wilesmith, M. S. Richards, P. Upton, and S. Johnston. 1995. The occurrence of *Mycobacterium bovis* infection in cattle in and

around an area subject to extensive badger (*Meles meles*) control. Epidemiol. Infect. **114**:179-193.

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moules, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. **393**:537-544.

Collins, C. H., J. M. Grange, and M. D. Yates. Tuberculosis bacteriology: organization and practice. 2nd ed. Oxford: Butterworth Heinemann, 1997, pp. 69-83.

Conaty, S. J., A. C. Hayward, A. Story, J. R. Glynn, F. A. Drobniewski, and J. M. Watson. 2004. Explaining risk factors for drug-resistant tuberculosis in England and Wales: contribution of primary and secondary drug resistance. Epidemiol. Infect. **132**:1099-1108.

Corbett, E. L., C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione, and C. Dye. 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch. Intern. Med. **163**:1009-1021.

Cosivi, O., J. M. Grange, C. J. Daborn, M. C. Raviglione, T. Fujikura, D. Cousins, R. A. Robinson, H. F. Huchzermeyer, K. de, I, and F. X. Meslin. 1998. Zoonotic tuberculosis due to *Mycobacterium bovis* in developing countries. Emerg. Infect. Dis. 4:59-70.

Cousins, D. V., R. A. Skuce, R. R. Kazwala, and J. D. Van Embden. 1998. Towards a standardized approach to DNA fingerprinting of *Mycobacterium bovis*. International Union Against Tuberculosis and Lung Disease, Tuberculosis in Animals Subsection. Int. J. Tuberc. Lung. Dis. **2**:471-478.

Cowan, L. S., L. Mosher, L. Diem, J. P. Massey, and J. T. Crawford. 2002. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. J. Clin. Microbiol. **40**:1592-1602.

Dale, J. W., R. M. Nor, S. Ramayah, T. H. Tang, and Z. F. Zainuddin. 1999. Molecular epidemiology of tuberculosis in Malaysia. J. Clin. Microbiol. **37**:1265-1268.

Dale, J. W., D. Brittain, A. A. Cataldi, D. Cousins, J. T. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. A. Skuce, C. Sola, D. van Soolingen, and V. Vincent. 2001. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. Int. J. Tuberc. Lung. Dis. **5**:216-219.

Dale, J. W., H. Al Ghusein, S. Al Hashmi, P. Butcher, A. L. Dickens, F. Drobniewski, K. J. Forbes, S. H. Gillespie, D. Lamprecht, T. D. McHugh, R. Pitman, N. Rastogi, A. T. Smith, C. Sola, and H. Yesilkaya. 2003. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS*6110*. J. Bacteriol. **185**:2555-2562.

Dale, J. W., G. H. Bothamley, F. Drobniewski, S. H. Gillespie, T. D. McHugh, and R. Pitman. 2005. Origins and properties of *Mycobacterium tuberculosis* isolates in London. J.Med. Microbiol. **54**:575-582.

Dalovisio, J. R., M. Stetter, and S. Mikota-Wells. 1992. *Rhinoceros' rhinorrhea*: cause of an outbreak of infection due to airborne *Mycobacterium bovis* in zookeepers. Clin. Infect. Dis. **15**:598-600.

de Boer, A. S., M. W. Borgdorff, P. E. de Haas, N. J. Nagelkerke, J. D. Van Embden, and D. van Soolingen. 1999. Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. J. Infect. Dis. **180** :1238-1244.

Delahay, R. J., C. L. Cheeseman, and R. S. Clifton-Hadley. 2001. Wildlife disease reservoirs: the epidemiology of *Mycobacterium bovis* infection in the European badger (*Meles meles*) and other British mammals. Tuberculosis (Edinb.). **81**:43-49.

Dingle, K. E., F. M. Colles, D. R. Wareing, R. Ure, A. J. Fox, F. E. Bolton, H. J. Bootsma, R. J. Willems, R. Urwin, and M. C. Maiden. 2001. Multilocus sequence typing system for *Campylobacter jejuni*. J. Clin. Microbiol. **39**:14-23.

Drobniewski, F., Y. Balabanova, M. Ruddy, L. Weldon, K. Jeltkova, T. Brown, N. Malomanova, E. Elizarova, A. Melentyey, E. Mutovkin, S. Zhakharova, and I. Fedorin. 2002. Rifampin- and multidrug-resistant tuberculosis in Russian civilians and prison inmates: dominance of the Beijing strain family. Emerg. Infect. Dis. **8**:1320-1326.

Drobniewski, F. A., A. Gibson, M. Ruddy, and M. D. Yates. 2003. Evaluation and Utilization as a public health tool of a national molecular epidemiological tuberculosis outbreak database within the United Kingdom from 1997 to 2001. J. Clin. Microbiol. **41**:1861-1868.

Durr, P. A., R. S. Clifton-Hadley, and R. G. Hewinson. 2000. Molecular epidemiology of bovine tuberculosis. II. Applications of genotyping. Rev. Sci. Tech. **19**:689-701.

Enright, M. C. and B. G. Spratt. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology. **144**:3049-3060.

Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes. 1998. IS*6110* transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. J. Bacteriol. **180**:2102-2109.

Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes. 1999a. IS*6110*-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. J. Bacteriol. **181**:1014-1020.

Fang, Z., C. Doig, A. Rayner, D. T. Kenna, B. Watt, and K. J. Forbes. 1999b. Molecular evidence for heterogeneity of the multiple-drug-resistant *Mycobacterium tuberculosis* population in Scotland (1990 to 1997). J. Clin. Microbiol. **37**:998-1003.

Ferdinand, S., G. Valetudie, C. Sola, and N. Rastogi. 2004. Data mining of *Mycobacterium tuberculosis* complex genotyping results using mycobacterial interspersed repetitive units validates the clonal structure of spoligotyping-defined families. Res. Microbiol. **155**:647-654.

Ferrazoli, L., M. Palaci, L. R. Marques, L. F. Jamal, J. B. Afiune, E. Chimara, M. C. Martins, M. A. Silva Telles, C. A. Oliveira, M. C. Palhares, D. T. Spada, and L. W. Riley. 2000. Transmission of tuberculosis in an endemic urban setting in Brazil. Int. J. Tuberc. Lung. Dis. **4**:18-25.

Filliol, I., S. Ferdinand, L. Negroni, C. Sola, and N. Rastogi. 2000. Molecular typing of *Mycobacterium tuberculosis* based on variable number of tandem DNA repeats used alone and in association with spoligotyping. J. Clin. Microbiol. **38**:2520-2524.

Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, G. Kallenius, E. Kassa-Kelembho, T.

Koivula, H. M. Ly, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. De Waard, C. Sola, and N. Rastogi. 2002. Global distribution of *Mycobacterium tuberculosis* spoligotypes. Emerg. Infect. Dis. **8**:1347-1349.

Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, M. L. Ho, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. De Waard, C. Sola, and N. Rastogi. 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. J. Clin. Microbiol. **41**:1963-1970.

Fomukong, N., M. Beggs, H. el Hajj, G. Templeton, K. Eisenach, and M. D. Cave. 1997. Differences in the prevalence of IS*6110* insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS*6110*. Tuber. Lung. Dis. **78**:109-116.

Frothingham, R. and W. A. Meeker-O'Connell. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology. **144**:1189-1196.

Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson. 2003. The complete genome sequence of *Mycobacterium bovis*. Proc. Natl. Acad. Sci. U.S.A **100**:7877-7882.

Gibson, A. L., G. Hewinson, T. Goodchild, B. Watt, A. Story, J. Inwald, and F. A. Drobniewski. 2004. Molecular epidemiology of disease due to *Mycobacterium bovis* in humans in the United Kingdom. J. Clin. Microbiol. **42**:431-434.

Githui, W. A., A. M. Jordaan, E. S. Juma, P. Kinyanjui, F. G. Karimi, J. Kimwomi, H. Meme, P. Mumbi, E. M. Streicher, R. Warren, P. D. van Helden, and T. C. Victor. 2004. Identification of MDR-TB Beijing/W and other *Mycobacterium tuberculosis* genotypes in Nairobi, Kenya. Int. J. Tuberc. Lung. Dis. **8**:352-360.

Glynn, J. R., J. Bauer, A. S. de Boer, M. W. Borgdorff, P. E. Fine, P. Godfrey-Faussett, and E. Vynnycky. 1999. Interpreting DNA fingerprint clusters of *Mycobacterium tuberculosis*. European Concerted Action on Molecular Epidemiology and Control of Tuberculosis. Int. J. Tuberc. Lung. Dis. **3**:1055-1060.

Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg. Infect. Dis. **8**:843-849.

Glynn, J. R., A. C. Crampin, M. D. Yates, H. Traore, F. D. Mwaungulu, B. M. Ngwira, R. Ndlovu, F. Drobniewski, and P. E. Fine. 2005a. The importance of recent infection with *Mycobacterium tuberculosis* in an area with high HIV prevalence: a long-term molecular epidemiological study in Northern Malawi. J. Infect. Dis. **192**:480-487.

Glynn, J. R., A. C. Crampin, H. Traore, M. D. Yates, F. D. Mwaungulu, B. M. Ngwire, D. D. Chaguluka, D. T. Mwafulirwa, S. Floyd, C. Murphy, F. A. Drobniewski, and P. E. Fine. 2005b. *Mycobacterium tuberculosis* Beijing genotype, northern Malawi. Emerg. Infect. Dis. **11**:150-3.

Godfrey-Faussett, P., P. Sonnenberg, S. C. Shearer, M. C. Bruce, C. Mee, L. Morris, and J. Murray. 2000. Tuberculosis control and molecular epidemiology in a South African gold-mining community. Lancet. **356**:1066-1071.

Goyal, M., N. A. Saunders, J. D. Van Embden, D. B. Young, and R. J. Shaw. 1997. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS*6110* restriction fragment length polymorphism. J. Clin. Microbiol. **35**:647-651.

Grange, J. M. and M. D. Yates. 1989. Incidence and nature of human tuberculosis due to *Mycobacterium africanum* in South-East England: 1977-87. Epidemiol. Infect. **103** :127-132.

Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser. 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms. Resolution of genetic relationships among closely related microbial strains. Genetics. **162**:1533-1543.

Haddad, N., A. Ostyn, C. Karoui, M. Masselot, M. F. Thorel, S. L. Hughes, J. Inwald, R. G. Hewinson, and B. Durand. 2001. Spoligotype diversity of *Mycobacterium bovis* strains isolated in France from 1979 to 2000. J. Clin. Microbiol. **39**:3623-3632.

Hardie, R. M. and J. M. Watson. 1992. *Mycobacterium bovis* in England and Wales: past, present and future. Epidemiol. Infect. **109**:23-33

Hawkey, P. M., E. G. Smith, J. T. Evans, P. Monk, G. Bryan, H. H. Mohamed, M. Bardhan, and R. N. Pugh. 2003. Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS*6110*-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. J. Clin. Microbiol. **41**:3514-3520.

Hayward, A. C. and J. M. Watson. 1995. Tuberculosis in England and Wales 1982-1993: notifications exceeded predictions. Commun. Dis. Rep. CDR Rev. **5**:R29-R33.

Heyderman, R. S., M. Goyal, P. Roberts, S. Ushewokunze, S. Zizhou, B. G. Marshall, R. Makombe, J. D. Van Embden, P. R. Mason, and R. J. Shaw. 1998.

Pulmonary tuberculosis in Harare, Zimbabwe: analysis by spoligotyping. Thorax. **53**:346-350.

Humphries, S. E., V. Gudnason, R. Whittall, and I. N. Day. 1997. Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholesterolemia. The IFCC Scientific Division: Committee on Molecular Biology Techniques. J. Int. Fed. Clin. Chem. **9**:156-161.

Hunter, P. R. and M. A. Gaston. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26**:2465-2466.

Kam, K. M., C. W. Yip, L. W. Tse, K. L. Wong, T. K. Lam, K. Kremer, B. K. Au, and D. van Soolingen. 2005. Utility of mycobacterial interspersed repetitive unit typing for differentiating multidrug-resistant *Mycobacterium tuberculosis* isolates of the Beijing family. J. Clin. Microbiol. **43**:306-313.

Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. **35**:907-914.

Kimerling, M. E., W. H. Benjamin, K. H. Lok, G. Curtis, and N. E. Dunlap. 1998. Restriction fragment length polymorphism screening of *Mycobacterium tuberculosis* isolates: population surveillance for targeting disease transmission in a community. Int. J. Tuberc. Lung. Dis. **2**:655-662.

Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. Van Embden. 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin. Microbiol. **37**:2607-2618.

Kumar, D., J. M. Watson, A. Charlett, S. Nicholas, and J. H. Darbyshire. 1997. Tuberculosis in England and Wales in 1993: results of a national survey. Public Health Laboratory Service/British Thoracic Society/Department of Health Collaborative Group. Thorax. **52**:1060-1067.

Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Sooligen, J. M. Musser, and B. N. Kreiswirth. 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non-random integration in the *dnaA-dnaN* region. Tuber. Lung. Dis. **79**:31-42.

Kwara, A., R. Schiro, L. S. Cowan, N. E. Hyslop, M. F. Wiser, H. S. Roahen, P. Kissinger, L. Diem, and J. T. Crawford. 2003. Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates. J. Clin. Microbiol. **41**:2683-2685.

Lockman, S., J. D. Sheppard, C. R. Braden, M. J. Mwasekaga, C. L. Woodley, T. A. Kenyon, N. J. Binkin, M. Steinman, F. Montsho, M. Kesupile-Reed, C. Hirschfeldt, M. Notha, T. Moeti, and J. W. Tappero. 2001. Molecular and

conventional epidemiology of *Mycobacterium tuberculosis* in Botswana: a population-based prospective study of 301 pulmonary tuberculosis patients. J. Clin. Microbiol. **39**:1042-1047.

Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen, and D. van Soolingen. 2003. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. Clin. Exp. Immunol. **133**:30-37.

Magdalena, J., P. Supply, and C. Locht. 1998. Specific differentiation between *Mycobacterium bovis* BCG and virulent strains of the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **36**:2471-2476.

Maguire, H., J. W. Dale, T. D. McHugh, P. D. Butcher, S. H. Gillespie, A. Costetsos, H. Al Ghusein, R. Holland, A. Dickens, L. Marston, P. Wilson, R. Pitman, D. Strachan, F. A. Drobniewski, and D. K. Banerjee. 2002. Molecular epidemiology of tuberculosis in London 1995-7 showing low rate of active transmission. Thorax. **57**:617-622.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. **95**:3140-3145.

Mandel, J. L. and V. Biancalana. 2004. Fragile X mental retardation syndrome: from pathogenesis to diagnostic issues. Growth. Horm. IGF. Res. **14 Suppl A**:S158-S165.

Mays, S., G. M. Taylor, A. J. Legge, D. B. Young, and G. Turner-Walker. 2001. Paleopathological and biomolecular study of tuberculosis in a medieval skeletal collection from England. Am. J. Phys. Anthropol. **114**:298-311.

Michalak, K., C. Austin, S. Diesel, M. J. Bacon, P. Zimmerman, and J. N. Maslow. 1998. *Mycobacterium tuberculosis* infection as a zoonotic disease: transmission between humans and elephants. Emerg. Infect. Dis. **4**:283-287.

Mokrousov, I., O. Narvskaya, E. Limeschenko, A. Vyazovaya, T. Otten, and B. Vyshnevskiy. 2004. Analysis of the allelic diversity of the mycobacterial interspersed repetitive units in *Mycobacterium tuberculosis* strains of the Beijing family: practical implications and evolutionary considerations. J. Clin. Microbiol. **42**:2438-2444.

Mostowy, S., J. Inwald, S. Gordon, C. Martin, R. Warren, K. Kremer, D. Cousins, and M. A. Behr. 2005. Revisiting evolution of *Mycobacterium bovis*. J. Bacteriol. 6386-6395.

Mostrom, P., M. Gordon, C. Sola, M. Ridell, and N. Rastogi. 2002. Methods used in the molecular epidemiology of tuberculosis. Clin. Microbiol. Infect. **8**:694-704.

Moxon E.R., P.B.Rainey, M.A.Nowark, and R.E.Lenski. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr. Biol. **4**:24-33.

Musser, J. M., A. Amin, and S. Ramaswamy. 2000. Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. Genetics. **155**:7-16.

Mwinga, A. and P. Bernard Fourie. 2004. Prospects for new tuberculosis treatment in Africa. Trop. Med. Int. Health. **9**:827-832.

Nerlich, A. G., C. J. Haas, A. Zink, U. Szeimies, and H. G. Hagedorn. 1997. Molecular evidence for tuberculosis in an ancient Egyptian mummy. Lancet. **350**:1404.

Nguyen, D., P. Brassard, D. Menzies, L. Thibert, R. Warren, S. Mostowy, and M. Behr. 2004. Genomic characterization of an endemic *Mycobacterium tuberculosis* strain: evolutionary and epidemiologic implications. J. Clin. Microbiol. **42**:2573-2580.

Niemann, S., E. Richter, and S. Rusch-Gerdes. 1999. Stability of *Mycobacterium tuberculosis* IS*6110* restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis. J. Clin. Microbiol. **37**:409-412.

Niemann, S., S. Rusch-Gerdes, M. L. Joloba, C. C. Whalen, D. Guwatudde, J. J. Ellner, K. Eisenach, N. Fumokong, J. L. Johnson, T. Aisu, R. D. Mugerwa, A. Okwera, and S. K. Schwander. 2002. *Mycobacterium africanum* subtype II is associated with two distinct genotypes and is a major cause of human tuberculosis in Kampala, Uganda. J. Clin. Microbiol. **40**:3398-3405.

Oevermann, A., G. E. Pfyffer, P. Zanolari, M. Meylan, and N. Robert. 2004. Generalized tuberculosis in llamas (*Lama glama*) due to *Mycobacterium microti*. J. Clin. Microbiol. **42**:1818-1821.

Ormerod, L. P., A. Charlett, C. Gilham, J. H. Darbyshire, and J. M. Watson. 1998. Geographical distribution of tuberculosis notifications in national surveys of England and Wales in 1988 and 1993: report of the Public Health Laboratory Service/British Thoracic Society/Department of Health Collaborative Group. Thorax. **53**:176-181.

Parish, T., D. A. Smith, G. Roberts, J. Betts, and N. G. Stoker. 2003. The *senX3-regX3* two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology. **149**:1423-1435.

Park, J. Y., L. Chen, A. Elahi, P. Lazarus, and M. S. Tockman. 2005. Genetic analysis of microsomal epoxide hydrolase gene and its association with lung cancer risk. Eur. J. Cancer. Prev. **14**:223-230.

Pongrapeeporn, K. S., Thepsuriyanon P., Leowattana W., Ong-Ajiyooth S., Kiartivich S., Yamwong P., Kasemsuk L., Kerdsaeng K., Nuchpramoll W., Laungsuwan S., and Amornrattana A. 2001. Inexpensive, rapid and convenient PCR-minigel SSCP protocol for polymorphisms and mutations analyses of LDL receptor gene. J. Med. Assoc. Thai. **84 Suppl 3**:S676-683.

Pretorius, G. S., P. D. van Helden, F. Sirgel, K. D. Eisenach, and T. C. Victor. 1995. Mutations in *katG* gene sequences in isoniazid-resistant clinical isolates of *Mycobacterium tuberculosis* are rare. Antimicrob.Agents Chemother. **39**:2276-2281.

Quint, A., I. Lerer, M. Sagi, and D. Abeliovich. 2005. Mutation spectrum in Jewish cystic fibrosis patients in Israel: implication to carrier screening. Am. J. Med. Genet. A. **136**:246-248.

Ratledge, C. and J.S. Stanford. The biology of mycobacteria. Volume 1. New York: Academic Press Inc, 1982, pp.198-199.

Reddy, P. H., M. Williams, and D. A. Tagle. 1999. Recent advances in understanding the pathogenesis of Huntington's disease. Trends. Neurosci. **22**:248-255.

Riley, R. L., C. C. Mills, W. Nyka, N. Weinstock, P. B. Storey, L. U. Sultan, M. C. Riley, and W. F. Wells. 1995. Aerial dissemination of pulmonary tuberculosis. A two-year study of contagion in a tuberculosis ward. 1959. Am. J. Epidemiol. **142**:3-14.

Roring, S., A. Scott, D. Brittain, I. Walker, G. Hewinson, S. Neill, and R. Skuce. 2002. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J. Clin. Microbiol. **40**:2126-2133.

Roring, S., A. N. Scott, H. R. Glyn, S. D. Neill, and R. A. Skuce. 2004. Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. Vet. Microbiol. **101**:65-73.

Rose, A. M., J. M. Watson, C. Graham, A. J. Nunn, F. Drobniewski, L. P. Ormerod, J. H. Darbyshire, and J. Leese. 2001. Tuberculosis at the end of the 20th century in England and Wales: results of a national survey in 1998. Thorax. **56**:173-179.

Ruddy, M. C., A. P. Davies, M. D. Yates, S. Yates, S. Balasegaram, Y. Drabu, B. Patel, S. Lozewicz, S. Sen, M. Bahl, E. James, M. Lipman, G. Duckworth, J. M. Watson, M. Piper, F. A. Drobniewski, and H. Maguire. 2004. Outbreak of isoniazid resistant tuberculosis in north London. Thorax. **59**:279-285.

Safi, H., J. Aznar, and J. C. Palomares. 1997. Molecular epidemiology of *Mycobacterium tuberculosis* strains isolated during a 3-year period (1993 to 1995) in Seville, Spain. J. Clin. Microbiol. **35**:2472-2476.

Salo, W. L., A. C. Aufderheide, J. Buikstra, and T. A. Holcomb. 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. Proc. Natl. Acad. Sci. U.S.A. **91**:2091-2094.

Savine, E., R. M. Warren, G. D. van der Spuy, N. Beyers, P. D. van Helden, C. Locht, and P. Supply. 2002. Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **40**:4561-4566.

Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston, G. F. Schecter, C. L. Daley, and G. K. Schoolnik. 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N. Engl. J. Med. **330**:1703-1709.

Smith, N. H., J. Dale, J. Inwald, S. Palmer, S. V. Gordon, R. G. Hewinson, and J. M. Smith. 2003. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. Proc. Natl. Acad. Sci. **100**:15271-15275.

Smith, R. M., F. Drobniewski, A. Gibson, J. D. Montague, M. N. Logan, D. Hunt, G. Hewinson, R. L. Salmon, and B. O'Neill. 2004. *Mycobacterium bovis* infection, United Kingdom. Emerg. Infect. Dis. **10**:539-541.

Sola, C., L. Horgen, K. S. Goh, and N. Rastogi. 1997. Molecular fingerprinting of *Mycobacterium tuberculosis* on a Caribbean island with IS*6110* and DRr probes. J. Clin. Microbiol. **35**:843-846.

Sola, C., I. Filliol, E. Legrand, I. Mokrousov, and N. Rastogi. 2001a. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS*1081*, IS*6110*, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J. Mol. Evol. **53**:680-689.

Sola, C., I. Filliol, M. C. Gutierrez, I. Mokrousov, V. Vincent, and N. Rastogi. 2001b. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. Emerg. Infect. Dis. **7**:390-396.

Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. U.S.A. **94**:9869-9874.

Supply, P., J. Magdalena, S. Himpens, and C. Locht. 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol. Microbiol. **26**:991-1003.

Supply, P., E. Mazars, S. Lesjean, V. Vincent, B. Gicquel, and C. Locht. 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol. Microbiol. **36**:762-771.

Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht. 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J. Clin. Microbiol. **39**:3563-3571.

Telenti, A., P. Imboden, F. Marchesi, T. Schmidheini, and T. Bodmer. 1993. Direct, automated detection of rifampin-resistant *Mycobacterium tuberculosis* by polymerase chain reaction and single-strand conformation polymorphism analysis. Antimicrob. Agents. Chemother. **37**:2054-2058.

Telenti, A., N. Honore, C. Bernasconi, J. March, A. Ortega, B. Heym, H. E. Takiff, and S. T. Cole. 1997. Genotypic assessment of isoniazid and rifampin

resistance in *Mycobacterium tuberculosis*: a blind study at reference laboratory level. J. Clin. Microbiol. **35**:719-723.

Torrea, G., C. Offredo, M. Simonet, B. Gicquel, P. Berche, and C. Pierre-Audigier. 1996. Evaluation of tuberculosis transmission in a community by 1 year of systematic typing of *Mycobacterium tuberculosis* clinical isolates. J. Clin. Microbiol. **34**:1043-1049.

van Belkum, A., S. Scherer, W. van Leeuwen, D. Willemse, L. van Alphen, and H. Verbrugh. 1997. Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. Infect. Immun. **65**:5017-5027.

van Deutekom, H., J. J. Gerritsen, D. van Soolingen, E. J. van Ameijden, J. D. Van Embden, and R. A. Coutinho . 1997. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. Clin. Infect. Dis. **25**:1071-1077.

Van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and . 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31**:406-409.

Van Embden, J. D., T. van Gorkom, K. Kremer, R. Jansen, B. A. Der Zeijst, and L. M. Schouls. 2000. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J. Bacteriol. **182**:2393-2401.

van Rie, A., R. M. Warren, N. Beyers, R. P. Gie, C. N. Classen, M. Richardson, S. L. Sampson, T. C. Victor, and P. D. van Helden. 1999. Transmission of a multidrug-resistant *Mycobacterium tuberculosis* strain resembling "strain W" among noninstitutionalized, human immunodeficiency virus-seronegative patients. J. Infect. Dis. **180**:1608-1615.

van Soolingen, D., P. E. de Haas, J. Haagsma, T. Eger, P. W. Hermans, V. Ritacco, A. Alito, and J. D. Van Embden. 1994. Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. J. Clin. Microbiol. **32**:2425-2433.

van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. Van Embden. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. **33**:3234-3238.

van Soolingen, D., T. Hoogenboezem, P. E. de Haas, P. W. Hermans, M. A. Koedam, K. S. Teppema, P. J. Brennan, G. S. Besra, F. Portaels, J. Top, L. M. Schouls, and J. D. Van Embden. 1997. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. Int. J. Syst. Bacteriol. **47**:1236-1245.

van Soolingen, D., M. W. Borgdorff, P. E. de Haas, M. M. Sebek, J. Veen, M. Dessens, K. Kremer, and J. D. Van Embden. 1999. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. J. Infect. Dis. **180**:726-736.

van Soolingen, D. 2001. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J. Intern. Med. **249**:1-26.

Warren, R. M., G. D. van der Spuy, M. Richardson, N. Beyers, M. W. Borgdorff, M. A. Behr, and P. D. van Helden. 2002. Calculation of the stability of the IS*6110* banding pattern in patients with persistent *Mycobacterium tuberculosis* disease. J. Clin. Microbiol. **40**:1705-1708.

Warren, R. M., E. M. Streicher, S. L. Sampson, G. D. van der Spuy, M. Richardson, D. Nguyen, M. A. Behr, T. C. Victor, and P. D. van Helden. 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. J. Clin. Microbiol. **40**:4457-4465.

Wilkinson, D., M. Pillay, J. Crump, C. Lombard, G. R. Davies, and A. W. Sturm. 1997. Molecular epidemiology and transmission dynamics of *Mycobacterium tuberculosis* in rural Africa. Trop. Med. Int. Health. **2**:747-753.

APPENDIX

APPENDIX 1

Spacer oligonucleotides sequences:

1. 5'-ATAGAGGGTCGCCGGTTCTGGATCA- 3'
2. 5'-CCTCATAATTGGGCGACAGCTTTG- 3'
3. 5'-CCGTGCTTCCAGTGATCGCCTTCTA- 3'
4. 5'-ACGTCATACGCCGACCAATCATCAG- 3'
5. 5'-TTTTCTGACCACTTGTGCGGGATTA- 3'
6. 5'-CGTCGTCATTTCCGGCTTCAATTTC- 3'
7. 5'-GAGGAGAGCGAGTACTCGGGGCTGC- 3'
8. 5'-CGTGAAACCGCCCCCAGCCTCGCCG- 3'
9. 5'-ACTCGGAATCCCATGTGCTGACAGC- 3'
10. 5'-TCGACACCCGCTCTAGTTGACTTCC- 3'
11. 5'-GTGAGCAACGGCGGCGGCAACCTGG- 3'
12. 5'-ATATCTGCTGCCCGCCCGGGGAGAT- 3'
13. 5'-GACCATCATTGCCATTCCCTCTCCC- 3'
14. 5'-GGTGTGATGCGGATGGTCGGCTCGG- 3'
15. 5'-CTTGAATAACGCGCAGTGGTCGGCTCGG- 3'
16. 5'-CGAGTTCCCGTCAGCGTCGTAAATC- 3'
17. 5'-GCGCCGGCCCGCGCGGATGACTCCG- 3'
18. 5'-CATGGACCCGGGCGAGCTGATCCTGGT- 3'
19. 5'-TAACTGGCTTGGCGCTGATCCTGGT- 3'
20. 5'-TTGACCTCGCCAGGAGAGAAGATCA- 3'
21. 5'-TCGATGTCGATGTCCCAATCGTCGA- 3'
22. 5'-ACCGCAGACGGCACGATTGAGACAA- 3'
23. 5'-AGCATCGCTGATGCGGTCCAGCTCG- 3'
24. 5'-CCGCCTGCTGGGTGAGACGTGCTCG- 3'
25. 5'-GATCAGCGACCACCGCACCCTGTCA- 3'
26. 5'-CTTCAGCACCACCATCATCCGGCGC- 3'
27. 5'-GGATTCGTGATCTCTTCCCGCGGAT- 3'
28. 5'-TGCCCCGGCGTTTAGCACACGACCA- 3'
29. 5'-AAATACAGGCTCCACGACACGACCA- 3'
30. 5'-GGTGCCCCGCGCCCTTTTCCAGCC- 3'
31. 5'-TCAGACACCTTCGCGTCGATCAAGT- 3'
32. 5'-GACCAAATAGGTATCGGCGTGTTCA- 3'
33. 5'-GACATGACGGCGGTGCCGCACTTGA- 3'
34. 5'-AAGTCACCTCGCCCACACCGTTCCAAC- 3'
35. 5'-TCCGTACGCTCGAAACGCTTCCAAC- 3'
36. 5'-CGAAATCCAGCACCACATCCGCAGC- 3'
37. 5'-CGCGAACTCGTCCACAGTCCCCCTT- 3'
38. 5'-CGTGGATGGCGGATGCGTTGTGCGC- 3'
39. 5'-GACGATGGCCAGTAAATCGGCGTGG- 3'
40. 5'-CGCCATCTGTGCCTCATACAGGTCC- 3'
41. 5'-GGAGCTTTCCGGCTTCTATCAGGTA- 3'
42. 5'-ATGGTGGGACATGGACGAGCGCGAC- 3'
43. 5'-CGCAGAATCGCACCGGGTGCGGGAG- 3'

APPENDIX 2

VNTR-ETR primers sequence

ETR A	5'- AAA TCG GTC CCA TCT CCT TCT TAT-3'
	5'- CGA AGC CTG GGG TGC CCG CGA TTT -3'
ETR B	5'- GCG AAC ACC AGG ACA GCA TCA TGG GTT -3'
	5'- GGC ATG CCG GTG ATC GAG TGG CTA TA -3'
ETR C	5'- GTG AGT CGC TGC AGA ACC TGC AG -3'
	5'- GGC GTC TTG ACC TAA TCG AGT G -3'
ETR D	5'- CAG GTC ACA ACG AGA GGA AGA GC -3'
	5'- GCG GAT CGG CCA GCG ACT CCT C -3'
ETR E	5'- CTT CGG CGT CGA AGA GAG CCT C -3'
	5'- CGG AAC GCT GGT CAC CAC CTA AG –3'

VNTR-MIRU-12 primers

MIRU 2	5'-CAG GTG CCC TAT CTG CTG ACG –3'
	5'- GTT GCG TCC GGC ATA CCA AC – 3'
MIRU 4	5'- GTC AAA CAG GTC ACA ACG AGA GGA A -3'
	5'- CCT CCA CAA TCA ACA CAC ACT GGT CAT -3'
MIRU 10	5'- ACC GTC TTA TCG GAC TGC ACT ATC AA-3'
	5'- CAC CTT GGT GAT CAG CTA CCT GAC CTA – 3'
MIRU 16	5' – CGG GTC CAG TCC AAC TAC CTC AAT-3'
	5'- GAT CCT CCT GAT TGC CCT GAC CTA-3'
MIRU 20	5'- CCC CTT CGA GTT AGT ATC GTC GGT T -3'
	5'- CCA TCA CCG TTA CAT CGA CGT CAT C -3'
MIRU 23	5'- CGA ATT CTT CGG TGG TCT CGA GT -3'
	5'- ACC GTC TGA CTC ATG GTG TCC AA -3'
MIRU 24	5'- GAA GGC TAT CCG TCG ATC GGT T -3'
	5'- GGG CGA GTT GAG CTC ACA GAA C -3'
MIRU 26	5'-GCG GAT AGG TCT ACC GTC GAA ATC -3'
	5'- TCC GGG TCT TAC AGC ATG ATC A – 3'
MIRU 27	5'- TCT GCT TGC CAG TAA GAG CCA -3'
	5'- GTG ATG GTG ACT TCG GTG CCT T -3
MIRU 31	5'- CGT CGA AGA GAG CCT CAT CAA TCA T -3'
	5'- AAC CTG ACC GAT GGC AAT ATC -3'
MIRU 39	5'- CGG TCA AGT TCA GCA CCT TCT ACA TC -3'
	5'- GCG TCC GTA CTT CCG GTT CAG -3'
MIRU 40	5'- GAT TCC AAC AAG ACG CAG ATA AAG A -3'
	5'- TCA GGT CTT TCT CTC ACG CTC TCG -3'

APPENDIX 3 - VNTR-MIRU-15 calling table

Copy No.	Capillary A						Copy No.	Capillary B						Copy No.	Capillary C		
	D2		D3		D4			D2		D3		D4			D2	D3	D4
	4	16	2	24	10	23		39	A	31	40	27	B		20	C	26
0	103	367	189	325	219	78	0	194	195	106	229	272	121	0	215	-	244
1	189	420	238	375	272	131	1	243	270	159	280	325	174	1	292	133	295
2	264	473	287	425	325	183	2	292	346	212	331	378	227	2	369	187	344
3	339	526	336	475	378	235	3	341	422	265	382	431	280	3	446	242	393
4	414	579	385	525	431	287	4	390	499	318	433	484	333	4	523	297	442
5	489	632	434	575	484	339	5	439	570	371	484	537	386	5	600	350	491
6	564		483	625	537	391	6	488	645	424	535	590	439	6	677	405	540
7	638		532		590	443	7	537	720	477	586	643	492	7		460	589
8	713		581		643	495	8	586	795	530	637		545	8		515	638
9			630			547	9	635	870	583			598	9		570	677
10						599	10			636			651	10		625	
11						651	11							11			

APPENDIX 4 - Univariate and multivariate analysis of basic epidemiological variables for clustering

		Non-clustered n = 1436	Clustered n = 372 (%)	Total n = 1808	p-value[a]	Odds ratio Unadjusted (95% CI)		Odds ratio Adjusted [b] (95% CI)	
Sex	Male	799	225 (22.0)	1024	0.209	1.00		1.00	
	Female	617	141 (18.6)	758		0.81	[0.64 – 1.03]	0.94	[0.72 – 1.23]
	Unknown	20	6 (23.1)	26					
Age	0 – 19	100	32 (24.2)	132	0.604	1.00		1.00	
	20 – 34	537	125 (18.9)	662		0.73	[0.47 – 1.13]	0.64	[0.39 – 1.05]
	35 – 39	445	119 (21.1)	564		0.84	[0.53 – 1.31]	0.69	[0.41 – 1.16]
	>= 60	335	92 (21.5)	427		0.86	[0.54 – 1.36]	0.61	[0.36 – 1.06]
	Unknown	19	4 (17.4)	23					
Ethnic Group	White	388	143 (26.9)	531	<.001	1.00		1.00	
	Black Caribbean	24	13 (35.1)	37		1.47	[0.73 – 2.96]	1.51	[0.67 – 3.39]
	Black African	226	67 (22.9)	293		0.80	[0.58 – 1.12]	0.71	[0.47 – 1.07]
	Black Other	9	6 (40.0)	15		1.81	[0.63 – 5.17]	1.20	[0.38 – 3.85]
	ISC	**522**	**85 (14.0)**	**607**		**0.44**	**[0.33 – 0.60]**	**0.45**	**[0.31 – 0.63]**
	Chinese	**42**	**3 (6.7)**	**45**		**0.19**	**[0.06 – 0.64]**	**0.21**	**[0.06 – 0.71]**
	Other	60	16 (21.1)	76		0.72	[0.40 – 1.30]	0.59	[0.31 – 1.16]
	Unknown	165	39 (19.1)	204					
Site of Disease	Extra-pulmonary	430	70 (14.0)	500	<.001	1.00		1.00	
	Pulm SSM+	522	155 (22.9)	677		1.82	[1.34 – 2.49]	1.41	[0.98 – 2.02]
	Pulm SSM-	**333**	**111 (25.0)**	**444**		**2.05**	**[1.47 – 2.85]**	**1.79**	**[1.23 – 2.61]**
	Unknown	151	36 (19.3)	187					
Previous Treatment	No	857	64 (6.9)	921	<.001	1.00		1.00	
	Yes	**78**	**22 (22.0)**	**100**		**3.78**	**[2.21 – 6.46]**	**3.74**	**[2.15 – 6.50]**
	Unknown	501	286 (36.3)	787					

Drug Susceptibility	Sensitive [c]	1356	340	(20.0)	1696	0.002	1.00		1.00	
	Resistant to at least 1 drug [c]	80	32	(28.6)	112		1.61	[1.11 – 2.35]	1.43	[0.93 – 2.19]
	Unknown	0	0		0					
UK Born [d]	No	784	187	(19.3)	971	0.011	1.00			
	Yes	384	128	(25.0)	512		1.40	[1.08 – 1.81]		
	Unknown	268	57	(17.5)	325					
Country of Birth [d]	UK	384	128	(25.0)	512	<.001	1.00			
	Europe	53	30	(36.1)	83		1.70	[1.04 – 2.77]		
	Africa	223	76	(25.4)	299		1.02	[0.74 – 1.42]		
	ISC	347	53	(13.3)	400		0.46	[0.32 – 0.65]		
	Asia	61	9	(12.9)	70		0.44	[0.21 – 0.92]		
	Other	15	9	(37.5)	24		1.80	[0.77 – 4.21]		
	Unknown	353	67	(16.0)	420					
Time since entry Into UK	UK born	384	128	(25.0)	512	0.045	1.00			
	>= 5 yrs	311	72	(18.8)	383		0.69	[0.50 – 0.96]		
	Between 1 & 4 yrs	221	54	(19.6)	275		0.73	[0.51 – 1.05]		
	< 1 yr	67	11	(14.1)	78		0.49	[0.25 – 0.96]		
	Unknown	453	107	(19.1)	560					

[a] χ-squared test or Fischer's Exact test
[b] Multivariate logistic regression adjusted for gender, age, ethnic group, previous treatment, site of disease, drug susceptibility
[c] Sensitive or resistant to Isoniazid, Ethambutol, Rifampicin, Pyrazinamide, Streptomycin
[d] Univariate analysis shown. These variables were combined to form Year of Entry into UK for multivariate analysis

ISC – Indian Sub - Continent
ssm+ - Sputum Smear Positive, ssm- - Sputum Smear Negative

APPENDIX 5 - Univariate and multivariate analysis of risk factors for clustering (case-control study)

		Non-clustered n = 315	Clustered n = 372 (%)	Total n = 687	p-value [a]	Odds ratio Unadjusted [95% CI]	Adjusted [b] [95% CI]
Homeless	No	156 (49.5)	221 (59.4)	377	<.001	1.00	1.00
	Yes	**2 (0.6)**	**21 (5.6)**	**23**		**7.41 [1.71 – 32.1]**	**5.47 [1.24 – 24.1]**
	Unknown	157 (49.8)	130 (34.9)	287			
Alcohol	No	149 (47.3)	211 (56.7)	360	<.001	1.00	
	Yes	**5 (1.6)**	**24 (6.5)**	**29**		**3.39 [1.26 – 9.09]**	
	Unknown	161 (51.1)	137 (36.8)	298			
Immunosuppressed	No	145 (46.0)	213 (57.3)	358	<.001	1.00	
	Yes	8 (2.5)	22 (5.9)	30		1.87 [0.81 – 4.32]	
	Unknown	162 (51.4)	137 (36.8)	299			
Prison	No	154 (48.9)	232 (62.4)	386	<.001	1.00	
	Yes	1 (0.3)	4 (1.1)	5		2.66 [0.29 – 24.0]	
	Unknown	160 (50.8)	136 (36.6)	296			
Drug injection	No	153 (48.6)	227 (61.0)	380	<.001	1.00	
	Yes	0 (0)	4 (1.1)	4		-	
	Unknown	162 (51.4)	141 (37.9)	303			
Outside UK	No	96 (30.5)	153 (41.1)	249	0.003	1.00	
	Yes	45 (14.3)	61 (16.4)	106		0.85 [0.54 – 1.35]	
	Unknown	174 (55.2)	158 (42.5)	332			
Compliance	Yes	138 (43.8)	181 (48.7)	319	0.003	1.00	
	No	17 (5.4)	41 (11.0)	58		1.84 [1.00 - 3.37]	
	Unknown	160 (50.8)	150 (40.3)	310			
Chest X-ray	Normal	26 (8.3)	35 (9.4)	61	<.001	1.00	
	Abnormal	109 (34.6)	179 (48.1)	288		1.22 [0.70 – 2.14]	

	Missing	180 (57.1)	158 (42.3)	338			
Cavities	No	57 (18.1)	80 (21.5)	137	0.002	1.00	
	Yes	30 (9.5)	65 (17.5)	95		1.16	[0.50 – 2.71]
	Missing	228 (72.4)	227 (61.0)	455			

[a] χ-squared test [b] Multivariate logistic regression adjusted for gender, age, ethnic group

APPENDIX 6 - Macroarray probes tried and tested

Batch	Gene(nt)	Probes (*wt*/*mt*)	Reason for change
1	*Oxy*R-37	*wt* gggcttcgcgcgttcgccgcgttttttttttt *mt* gggtttcgcgcgttcgccgcgttttttttttt	Weak hybridisation signal/ No hybridisation signal
	KatG-87	*wt* catatgaaataccccgtcgagttttttttttt *mt* catatgaaatacccagtcgagttttttttttt	No hybridisation signal
	RpoB-2646	*wt* ctccgacggtgacaagctggctttttttttt *mt* ctccgacgggacaagctggctttttttttt	Hybridisation to *wt* and *mt* probes
	RpoB-3243	*wt* gcctacggtgctgcctacatttttttttt *mt* gcctacggtgccgcctacatttttttttt	No hybridisation signal
2	*Oxy*R-37	*wt* ccaccgcggcgaacgcgcgaagccctttttttttt *mt* ccaccgcggcgaacgcgcgaaacccttttttttttt	Hybridisation to *wt* and *mt* probes
	KatG-87	*wt* ccaccgcggcgaacgcgcgaagccctttttttttt *mt* ccaccgcggcgaacgcgcgaaacccttttttttttt	Weak hybridisation signal
	RpoB-2646	*wt* tgccggccggccagcttgtcaccgtttttttttt *mt* tgccggccggccagcttgtccccgtttttttttt	Weak hybridisation signal
	RpoB-3243	*wt* cagctcctgcagggtgtaggcagcatttttttttt *mt* cagctcctgcagggtgtaggcggcatttttttttt	Weak hybridisation signal
3	*RpoB*-2646	*wt* tgccggccggccagcttgtcaccgtttttttttt *mt* tgccggccggccagcttgtccccgtttttttttt	Weak hybridisation signal
	RpoB-3243	*wt* ggcagcaccgtaggcctgcatggctttttttttttttttttttt *mt* ggcggcaccgtaggcctgcatggctttttttttttttttttttt	Weak hybridisation signal
4	*Oxy*R-37	*wt* Ccaccgcggcgaacgcgcgaagccctttttttttt *mt* ccaccgcggcgaacgcgcgaaacccttttttttttt	Weak hybridisation signal.

	KatG-87	*wt* accccgtcgagggcggcggaaaccagttttttttt *mt* acccagtcgagggcggcggaaaccagttttttttt	Weak hybridisation signal
	RpoB-2646	*wt* tgccggccggccagcttgtcaccgttttttttt *mt* tgccggccggccagcttgtccccgttttttttt	Weak hybridisation signal
	RpoB-3243	*wt* cagctcctgcagggtgtaggcagcatttttttttt *mt* cagctcctgcagggtgtaggcggcatttttttttt	Weak hybridisation signal
5	*OxyR*-37	*wt* Ccaccgcggcgaacgcgcgaagccctttttttttttttttttttt	Mutant isolate hybridised to *wt* probe
	RpoB-3243	*wt* tcctgcagggtgtaggcagcaccgtttttttttttttttttttt	Weak hybridisation
6	*RpoB*-3243	*wt* ctcctgcagggtgtaggagctttttttttttttttttttt	Mutant isolate hybridised to *wt* probe
7	*OxyR*-37	*wt* ccacctcggcgaacggcgcgaagcctttttttttttttttttttt	Mutant isolate hybridised to *wt* probe
	RpoB-3243	*wt* ggcctgccagccagcttgtcaccgtttttttttttttttttttt	Mutant isolate hybridised to *wt* probe
8	*RpoB*-3243	*wt* cggccagcttggcaccgtcggtttttttttttttttttttt *wt* gcttgtcaccgtcggagatcttgtttttttttttttttttttt	Weak hybridisation

JOURNAL OF CLINICAL MICROBIOLOGY, Oct. 2004, p. 4536–4544 Vol. 42, No. 10
0095-1137/04/$08.00+0 DOI: 10.1128/JCM.42.10.4536–4544.2004

High Rates of Clustering of Strains Causing Tuberculosis in Harare, Zimbabwe: a Molecular Epidemiological Study

Philippa J. Easterbrook,[1]* Andrea Gibson,[2] Shahed Murad,[1] Dianie Lamprecht,[2] Natalie Ives,[1] Alex Ferguson,[3] Odette Lowe,[4] Peter Mason,[3] Angelika Ndudzo,[3] Alfred Taziwa,[3] Robert Makombe,[5] Lovemore Mbengeranwa,[5] Christophe Sola,[6] Nalim Rostogi,[6] and Francis Drobniewski[2]

Department of HIV/GUM, The Guy's, King's and St. Thomas' School of Medicine, King's College Hospital, Denmark Hill Campus,[1] and Health Protection Agency, Mycobacterial Reference Unit, Guy's, King's and St. Thomas' School of Medicine, King's College Hospital (Dulwich),[2] London, United Kingdom; University of Zimbabwe[3] and Harare City Health Department,[5] Harare, and National TB Reference Laboratory, Bulawayo,[4] Zimbabwe; and Institute Pasteur, Pointe à Pitre, Guadeloupe[6]

Received 13 November 2003/Returned for modification 5 February 2004/Accepted 18 June 2004

We examined the pattern of tuberculosis (TB) transmission (i.e., reactivation versus recent transmission) and the impact of human immunodeficiency virus (HIV) infection in Harare, Zimbabwe. Consecutive adult smear-positive pulmonary TB patients presenting to an urban hospital in Harare were enrolled. A detailed epidemiological questionnaire was completed, and tests for HIV type 1 and CD4 cell counts were performed for each patient. Molecular fingerprinting of the genomic DNA recovered from cultures of sputum was performed by two molecular typing methods: spacer oligonucleotide typing (spoligotyping) and analysis of variable number of tandem DNA repeats (VNTRs). A cluster was defined as isolates from two or more patients that shared the same spoligotype pattern or the same VNTR pattern, or both. DNA suitable for typing was recovered from 224 patients. The prevalence of HIV infection was 79%. Of 187 patient isolates (78.6%) typed by both spoligotyping and analysis of VNTRs, 147 were identified as part of a cluster by both methods. By spoligotyping alone, 84.1% of patient isolates were grouped into 20 clusters. The cluster size was generally <8 patient isolates, although three large clusters comprised 68, 25, and 23 patient isolates. A total of 89.4% of the patient isolates grouped into 12 clusters defined by analysis of VNTRs, with 2 large clusters consisting of 127 and 13 patient isolates, respectively. Thirty-six percent of patient isolates with a shared spoligotype and 17% with a shared VNTR pattern were geographically linked within Harare, but they were not linked on the basis of the patient's home district. In a multivariate analysis, there were no independent predictors of clustering, including HIV infection status. Comparison with the International Spoligotype database (Pasteur Institute, Pointe à Pitre, Guadeloupe) demonstrated that our three largest spoligotype clusters are well recognized and ubiquitous in Africa. In this epidemiologically well characterized urban population with a high prevalence of HIV infection, we identified a very high level of strain clustering, indicating substantial ongoing recent TB transmission. Geographic linkage could be detected in a proportion of these clusters. A small group of actively circulating strains accounted for most of the cases of TB transmission.

Tuberculosis (TB) has reemerged as an important public health problem in Zimbabwe. Following a stabilization of the TB rates in the early 1980s at approximately 50 per 100,000 population, since 1987 there has been a dramatic increase to 250 per 100,000 population, which is largely attributable to the human immunodeficiency virus (HIV) epidemic in sub-Saharan Africa (9, 28).

Strain-specific markers for the differentiation of *Mycobacterium tuberculosis* complex strains are reliable tools for the identification of specific strains and are useful for epidemiological studies of TB. The most extensively used differentiation method is restriction fragment length polymorphism (RFLP) typing, which uses the insertion sequence IS*6110* to differentiate clinical isolates (39). However, this technique is technically demanding and expensive and requires viable bacterial growth, and so it may not be optimal for large epidemiological studies in resource-poor settings. New PCR-based typing methods such as spacer oligonucleotide typing (spoligotyping), based on polymorphisms in the direct repeat locus, require less DNA and have been increasingly used for the rapid typing of isolates. However, spoligotyping used alone is less discriminatory than RFLP typing and tends to overestimate the number of epidemiological links, so the use of a second method for confirmation of the results is recommended (22). Fingerprinting based on the variable number of tandem DNA repeats (VNTRs) is a further highly reproducible rapid typing method (15) that yields a high level of cluster discrimination comparable to that of RFLP analysis when it is used as a second-line test with spoligotyping (13, 26).

Use of conventional epidemiological data in combination with these DNA techniques can help answer important questions about the epidemiology of TB in large populations by distinguishing between newly acquired and reactivated disease (1, 4, 6, 7, 16, 23, 32, 37). Patients whose isolates have identical

* Corresponding author. Mailing address: Department of HIV/GU Medicine, The Guy's, King's and St. Thomas's School of Medicine, Weston Education Centre, Cutcombe Rd., King's College Hospital, Denmark Hill Campus, London SE5 9RT, United Kingdom. Phone: 020 7848 5770. Fax: 020 7848 5769. E-mail: philippa.easterbrook@kcl.ac.uk.

patterns (i.e., a cluster) are likely to have been infected recently and can be targeted for epidemiological investigation to identify a chain of transmission, whereas patients whose isolates demonstrate unique patterns are likely to have reactivation of a latent infection. Most previous studies that have used these DNA fingerprinting techniques have focused on the tracing of TB outbreaks within a hospital or other high-risk environments and the dissemination of multidrug-resistant strains, in which isolates from epidemiologically related patients are usually identical (8, 10, 11, 14). Although it was previously thought that 90% of TB cases in the developed world resulted from the endogenous reactivation of latent infection, population-based RFLP studies conducted in the United States and Western Europe, areas with a relatively low incidence of TB, show that recent infection accounts for up to half of the cases among both HIV-infected and HIV-negative patients in urban areas (1, 4, 6, 7, 16, 23, 32, 37). These studies also showed that only a fraction of the main TB transmission routes are disclosed by classical contact-tracing practices.

In Africa, where TB is endemic, it has generally been considered that most cases of TB in HIV-infected and HIV-negative patients result from reactivation (9). However, there are few data to support this, as few detailed molecular epidemiological studies have been undertaken in countries with high rates of both TB and HIV infection. It has been shown that despite the more limited strain diversity in sub-Saharan Africa, there is still adequate strain variability among *M. tuberculosis* isolates from patients living in developing countries to permit the application of molecular approaches to the tracking of TB transmission (24, 25). Our objectives were, first, to examine the pattern of TB transmission (i.e., the relative frequency of reactivation versus that of recent infection) and the role of HIV infection using two molecular typing methods, spoligotyping and analysis of VNTRs, in tandem with epidemiological data and, second, to identify the risk factors for recent transmission in a well-characterized cohort of smear-positive TB patients in Harare, Zimbabwe.

(This study was presented in part at the 9th Conference on Retroviruses and Opportunistic Infections, Seattle, Wash., 24 to 28 February 2002 [P. Easterbrook, A. Gibson, S. Murad, A. Ferguson, P. Mason, A. Ndudza, L. Mbengeranwa, and F. Drobniewski, Abstr. 9th Conf. Retrovir. Opportunistic Infect., abstract 621 W, 2002].)

MATERIALS AND METHODS

A total of 516 consecutive adult smear-positive pulmonary TB patients presenting to the Beatrice Road TB Hospital in Harare, Zimbabwe, between May and October 1997 were enrolled in the study. The Beatrice Road TB Hospital is the main referral center for TB and other infectious diseases in Harare and captures the majority of TB cases in the city, which has a population of about 1.25 million. Data for each patient were collected by using a standardized questionnaire and pro forma by a trained research nurse and included demographic data (age, gender, occupation, marital status, number of children, rural home district, present address and all previous residence addresses over the last year in Harare, household size, number of rooms and household crowding level [i.e., number of dwellers per number of rooms], medical history [past history of TB and treatment, *M. bovis* BCG vaccination, and history of other medical conditions, e.g., diabetes], the history of TB in the household within the last 5 years, and smoking and alcohol consumption). A chest X ray was obtained for each patient, and a comprehensive physical examination focused on the detection of clinical features suggestive of HIV disease. Informed consent to perform HIV testing was obtained from all patients, and the study was approved by the Medical Research Council Ethics Committee of Zimbabwe.

Laboratory methods. HIV infection status was determined by an enzyme-linked immunosorbent assay (Dupont, Wilmington, Del.). CD4 cell counts were measured by flow cytometry (FACScan; Becton Dickinson, Paramus, N.J.). Three smear-positive sputum samples from each patient were sent to the TB reference laboratory in Bulawayo, Zimbabwe, by the usual transport system and were cultured on Lowenstein-Jensen medium. Cultures and drug susceptibility testing were performed for all 502 patients with microbiologically confirmed cases of TB. Isolates were categorized into drug-sensitive and single-drug- or multidrug-resistant strains.

Molecular typing was performed at the National Mycobacterial Reference Unit, Health Protection Agency, London, United Kingdom, after transport of the cultures by overnight courier from Zimbabwe. Spoligotyping was performed with genomic DNA extracted from cultures of sputum by a standard phenol-chloroform method (43). The spacers between the direct repeats in the target region were amplified by using two 18-nucleotide primers (primer 5′-CCAAGA GGGGACGGAAAC-3′ and biotinylated primer 5′-GGTTTTGGGTCTGACG AC-3′). The PCR products were then hybridized to a Biodyne C membrane (Isogen Bioscience, Maarsen, The Netherlands). This membrane contains immobilized synthetic oligomeric spacer sequences derived from the direct-repeat region of *M. tuberculosis* H37Rv and *M. bovis* BCG. Hybridized DNA was detected by using an enhanced chemiluminescence kit (Amersham International plc, Little Chalfont, United Kingdom), with exposure to X-ray film producing a pattern or profile reminiscent of a bar code.

Analysis of VNTRs was performed as described previously (15), but with the following modifications: DNA was extracted from the cultures as described above. PCR was performed in a total volume of 20 μl containing 11 μl of HotStarTaq DNA polymerase (Qiagen, Chichester, United Kingdom) and 10 to 100 ng of DNA sample. An initial denaturation at 94°C for 15 min was followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 1 min, and extension at 72°C for 2 min, followed by a final extension at 72°C for 10 min. An aliquot (10 μl) from each reaction tube was run on a 2% Nusieve GTG Agarose gel (Flowgen, Leicestershire, United Kingdom). Molecular weight standards (100-bp ladder [Promega, Southampton, United Kingdom] and a 20-bp ladder [Sigma-Aldrich, Dorset, United Kingdom]) were loaded every eight lanes. The gel was stained with ethidium bromide and visualized under UV light. Once the size of the PCR product was determined, the number of copies for each exact tandem repeats (ETR) (A to E) was determined by a previously described method (15) and a five-digit number representing the allele profiles was created.

Statistical analysis. Patient isolates were categorized into one of two groups according to whether the spoligotype or VNTR pattern was unique or was identical to that for another sample (i.e., a cluster). A cluster was defined as a group of two or more patient isolates that shared the same spoligotype or VNTR pattern. Clusters were considered false positive if the sample was processed in the microbiology laboratory on the same day that a specimen with a positive smear from another patient with the same typing pattern was processed. We also reanalyzed the data by assuming that each cluster of n patients comprised one source case and that all the other cases in the clusters (i.e., $n - 1$) were due to recently acquired disease through active transmission (32). Therefore, a minimum estimate of the proportion of cases of TB caused by recent transmission was calculated as (number of clustered patients − number of clusters)/total number of patients. Patient isolates contained within clusters and those with unique patterns were compared by chi-square tests or Fisher's exact test and by Mann-Whitney U tests for categorical and continuous variables, respectively. Risk factors for clustering identified by univariate analysis ($P < 0.2$) were then included in a multivariate logistic regression model to identify independent risk factors for clustering, with clustered and nonclustered being the dependent outcomes and with a cluster defined as two or more patient isolates that shared the same (i) spoligotype and VNTR patterns, (ii) the same spoligotype pattern, and (iii) the same VNTR pattern. Since the overall rate of clustering may not always be informative, we also examined whether certain Harare districts and rural home districts were associated with a higher frequency of overall clustering and with particular clusters identified by spoligotyping or analysis of VNTRs using a binomial probability test and one-sided P values after exclusion of those districts in which only one study patient was recorded and exclusion of those patients within a district who were part of a cluster but from another district. For each patient within a cluster, the location of the patient's home was recorded by using the Harare Street Atlas. Individuals living within 0.5 km of a patient who was part of the same cluster were considered part of a geographically based cluster. Data were analyzed by using STATA software (version 7.0; STATA Corporation, College Station, Tex.). Taxatron software (Institut Pasteur, Paris,

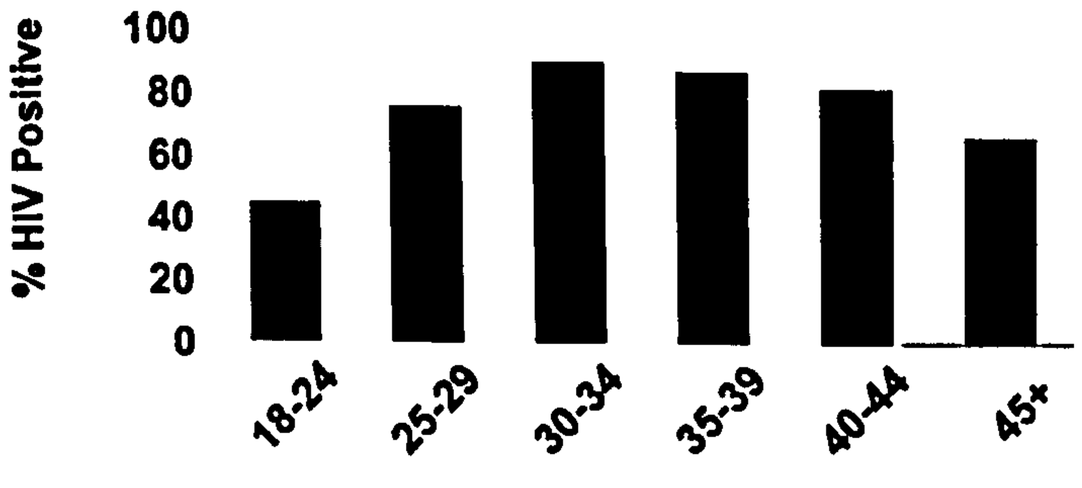


FIG. 1. Prevalence of HIV-1 infection among TB patients by age group.

France) was used to calculate the molecular weights of the hybridizing bands and to compare the isolates with those in the International Spoligotype database.

RESULTS

Study population. A total of 516 patients with smear-positive TB who presented to the Beatrice Road TB Hospital from May to October 1997 were enrolled in the study. Complete data were available for 502 patients, and these patients were included in the analysis. Of the 502 patients, 371 (74%) were HIV positive and 65% were male. The median age at presentation was 31 years (interquartile range [IQR], 27 to 40 years). Among the 502 patients, 204 (41%) gave a history of recent household contact with TB and 44 (9%) had a past history of TB. There was a striking difference in the prevalence of HIV infection by age (Fig. 1). For the HIV-positive patients, the median CD4 cell count at the time of presentation was 150 × 10^6/liter (IQR, 64 to 291/liter). Drug resistance was identified in isolates from only 17 (4%) of 458 patients.

Spoligotyping and VNTR patterns. Useable DNA was recovered from 224 of the 228 individual cultures available for 502 (44.6%) patients. A significant number of the remaining cultures were extensively contaminated with bacteria or yeasts, which led to degradation of the medium. Spoligotyping was performed on 214 samples (after the exclusion of 10 contaminated samples), and analysis of VNTRs was performed on 198 samples (after the exclusion of 3 contaminated samples). There were no statistically significant differences in the characteristics of the 187 patients from whom DNA was available for typing compared to those of the 317 patients from whom DNA was not available. Of the 187 patients with available spoligotype and VNTR typing data, 147 (78.6%) were identified as part of a cluster on the basis of both methods, with an additional 20 identified as part of a cluster on the basis of analysis of VNTRs alone and 12 identified as a cluster on the basis of spoligotyping alone.

By using spoligotyping alone, 54 distinct spoligotypes were identified and a total of 180 patient isolates (84.1%) were grouped into 20 clusters (or 15 clusters on the basis of the $n - 1$ approach, as 10 patient isolates were in clusters with only 2 isolates). The cluster size was generally less than 8 isolates (Table 1); however, three large clusters comprised 68, 25, and 23 patient isolates, respectively, and these accounted for 50% of the isolates. In comparison, there were 33 distinct strains by analysis of VNTRs, and 177 patient isolates (89.4%) grouped in 12 clusters defined by analysis of VNTRs (or 10 clusters on the basis of the $n - 1$ approach, as 4 patient isolates were in clusters with only 2 isolates), with 2 large clusters comprising 127 isolates (which accounted for 64.1% of the isolates in the study) and 13 patients. The minimum estimate for the proportion of TB cases due to recent transmission was 74.8% [(180 − 20)/214, i.e., (number of clustered patients − number of clusters)/total number of patients] for spoligotyping and 83.3% [(177 − 12)/198] for analysis of VNTRs.

Analysis of VNTRs was less discriminatory as a first-line fingerprinting method (Table 1). Eleven of the 12 (91.7%) clusters identified by analysis of VNTRs could be further subdivided by spoligotyping into distinct subtypes, with each cluster containing from 2 to 23 subtypes. Conversely, only 8 of 20 (40%) clusters obtained by spoligotyping were further subdivided by analysis of VNTRs. Within the three largest clusters obtained by spoligotyping, analysis of VNTRs identified 8, 11,

TABLE 1. Clustering results obtained by analysis of VNTRs as a first-line test followed by subclustering by spoligotyping and obtained by spoligotyping as a first-line test followed by subclustering by analysis of VNTRs

Primary typing method and cluster size (no. of isolates)	Allele designation by analysis of VNTRs	No. of subclusters by:	
		Spoligotyping	Analysis of VNTRs[a]
Analysis of VNTRs			
5	3233*3	3	
4	2123*2	2	
4	2123*3	2	
3	2143*1	2	
5	2143*2	4	
127	2143*3	23	
2	4243*5		
13	3243*3	6	
2	2243*2	2	
6	2243*3	3	
3	2343*3	2	
3	3233*2	3	
Spoligotyping			
4			—[b] (NT = 3)
2			— (NT=1)
68			8 (NT=7)
2			— (NT=4)
23			4 (NT=3)
3			2
3			— (NT=2)
5			2
8			—
4			2 (NT=1)
2			2
2			—
3			—
2			— (NT=1)
7			—
25			11 (NT=3)
4			—
3			2
7			
3			

[a] Expressions in parentheses represent numbers of samples not tested (NT) because a strain type was not available by analysis of VNTRs for some isolates analyzed by spoligotyping.

[b] —, no subclusters identified.

TABLE 2. Characteristics of clustered and nonclustered patients determined by spoligotyping[a]

Characteristic	Value for group		*P* value
	Cluster of ≥2 isolates (n = 180)	Noncluster (n = 34)	
Median age (yr [IQR])	30 (26–38.7)	31 (26.7–37.2)	0.71
No. male (%)	110 (61.1)	28 (82.3)	0.02
No. (%) HIV-1 seropositive	138 (77.1)	30 (88.2)	0.17
No. (%) resistant with isolates to one or more anti-TB drug	4 (2.2)	3 (9.1)	0.08
No. (%) with household crowding level of:[b]			0.30
<1	19 (10.8)	4 (12.1)	
1–2	29 (16.5)	9 (27.3)	
>2	128 (72.7)	20 (60.6)	
No. (%) who ever smoked	59 (32.8)	20 (58.8)	0.004
No. (%) with:			
BCG scar	161 (89.9)	32 (94.1)	0.75
TB in household in last 5 year	65 (36.7)	12 (35.3)	0.89
Past history of TB	11 (6.1)	5 (14.7)	0.08
Symptoms of HIV infection[c]			0.70
1 symptom	53 (29.6)	12 (35.3)	
≥2 symptoms	44 (24.6)	9 (26.5)	
Harare districts[d]			0.47
Median CD4 count (10^6/liter [IQR])	232.5 (87–443)	262.5 (125–396)	0.83

[a] Comparison of clusters defined by two or more isolates by analysis of VNTRs (n = 177) and noncluster (n = 21) and those defined as a cluster by both spoligotyping and analysis of VNTRs (n = 147) versus noncluster (n = 8) yielded similar results, except that gender was no longer statistically significant.
[b] Calculated from number of dwellers/number of rooms in household.
[c] Chronic diarrhea for >1 month, recurrent pneumonia, wasting syndrome, herpes zoster rash, and persistent genital ulcers.
[d] Eleven districts: Budiriro, Chitungwiza, Dzivaresekwa, Glen Norah, Glen View, Highfield, Kambuzuma, Mbara, Mufakose, Mabunku, and Kuwadzana.

and 4 additional unique strains, respectively. However, a corresponding fingerprint obtained by analysis of VNTRs was not available for five distinct spoligotypes, so the discriminatory power of analysis of VNTRs may have been underestimated.

Risk factors for clustering. To identify risk factors for recent infection with *M. tuberculosis*, the 180 patients whose isolates were in spoligotype-defined clusters were compared with the 34 patients whose isolates were not in clusters. Univariate analysis (Table 2) showed that patients whose isolates were in clusters were less likely to be male (P = 0.02), less likely to harbor a drug-resistant isolate (P = 0.08), and less likely to have a history of TB (P = 0.08) but were more likely to have never smoked (P = 0.004) than patients whose isolates had unique DNA patterns. Factors not associated with clustering were age, HIV infection status, household crowding level, history of TB in the household in the last 5 years, number of symptoms suggestive of HIV (i.e., chronic diarrhea of >1 month, recurrent pneumonia, wasting syndrome, herpes zoster rash, and persistent genital ulcers), presence of a BCG scar, and CD4 cell count. We found similar results when we compared the 177 patients whose isolates were in clusters identified by analysis of VNTRs and the 21 patients whose isolates were not clustered and when we compared the 147 patients whose isolates were defined as part of a cluster on the basis of both spoligotyping and analysis of VNTRs and 8 patients whose isolates were not clustered, except that gender was no longer statistically significant in these analyses. In a multivariate analysis, we found no independent predictors of clustering. We also performed a similar analysis to establish risk factors for membership in a large cluster (≥6, ≥12, and ≥20 patients with a shared strain). We found that a history of TB in the household within the last 5 years was significantly less common among those within a cluster comprising six or more patient isolates with a shared spoligotype and VNTR pattern (30.2%) than in patients whose isolates comprised a cluster with less than six isolates (60.7%; P = 0.003), and the CD4 cell count was significantly higher in the former group than in the latter group (320×10^6 and 240×10^6 cells/liter, respectively; P = 0.02). In a multivariate analysis, a history of TB in the household in the last 5 years (odd ratio = 0.24; 95% confidence interval = 0.09 to 0.68; P = 0.007) and infection with a drug-resistant isolate (odds ratio = 0.05; 95% confidence interval = 0.003 to 0.95; P = 0.05) were inversely associated with the presence of clustering.

Geographic mapping of clusters identified by spoligotyping and analysis of VNTRs. Overall, we found no evidence for a significantly higher frequency of clustering within certain Harare districts of residence or on the basis of the patients' rural home districts in Zimbabwe. However, when we repeated the analysis according to the individual clusters identified by spoligotyping and analysis of VNTRs, 7 of 20 (35%) strains that clustered by spoligotyping and 2 of 12 (16.7%) strains that clustered by analysis of VNTRs had a higher frequency in certain Harare districts (but not the rural home districts) com-

TABLE 3. Geographic clusters within Harare districts for 7 of 20 clusters identified by spoligotyping and 2 of 12 clusters identified by analysis of VNTRs

Typing method and cluster code	No. of patients in cluster	Frequency (%) in overall study population	Harare district	No. of study patients in cluster in district/no. of study patients in district	Frequency (%) of cluster isolate in district	*P* value
Spoligotyping						
777777606060771	68	37.8	Glen View	10/19	52.6	0.14
			Highfield	13/23	56.5	0.06
			Kuwadzana	4/7	57.1	0.25
			Mbare	12/26	46.1	0.25
			Waterfalls	2/2	100	0.14
777777607760771	23	12.8	Chitungwiza	2/5	40	0.13
			Dzivaresekwa	2/5	40	0.13
			Kambuzuma	2/3	66.7	0.05
777777677760771	5	2.8	Glen View	2/19	10.5	0.11
777777604060731	7	3.9	Mufukose	3/13	23.1	0.01
777777777760731	25	13.9	Mufakose	4/13	30.8	0.09
777777606060631	4	2.2	Highfield	2/23	9.0	0.09
777777606060731	7	3.9	Braeside	2/11	18.2	0.07
Analysis of VNTRs						
2123*2	4	2.0	Highfield	2/26	7.7	0.10
			Glen View	2/22	9.1	0.07
2143*3	127	64.1	Arcadia	2/2	100	0.41
			Budiriro	8/9	88.9	0.11
			Glen View	16/22	72.7	0.26
			Harare	2/2	100	0.41
			Kambuzuma	3/4	75	0.54
			Mabvuku	6/7	85.7	0.22
			Southerton	3/3	100	0.26

pared to the frequency of the strain cluster in the overall study population, although the differences did not generally attain statistical significance (Table 3). Isolates in one large cluster identified by spoligotyping (n = 68), which accounted for 37.8% of all isolates, were found at a higher frequency (>50%) in five districts, while a further cluster of 23 isolates (12.8% of the study population) was found among >40% of patients in three districts. Five other smaller clusters identified by spoligotyping were also disproportionately represented in certain Harare districts. A similar picture emerged with 2 of the 12 clusters identified by analysis of VNTRs. The largest cluster identified by analysis of VNTRs (n = 127), which accounted for 64.1% of all isolates, was found at a higher frequency (>80%) in five districts, although this did not attain statistical significance. A further small cluster of four patients identified by analysis of VNTRs was found among patients from only two districts. When we mapped the locations of the patients' homes on a map of Harare, we found that the vast majority of the patients in patient clusters within a particular district resided within 0.5 km of each other.

Relationships between isolates identified by spoligotyping (Table 4). Linkage with the International Spoligotype database (Pasteur Institute, Pointe à Pitre, Guadeloupe) demonstrated that our three largest spoligotype cluster types (International Spoligotype Database code numbers 59, 53, and 42) are all well recognized and ubiquitous in Africa. All clusters between types 804 and 815 have not yet been described elsewhere and could be considered a result of the specific evolution of type 59. Type 1 is the Beijing/W strain, which was represented by the isolates from four patients in the study population, none of whom were geographically linked, although two of the patients originated from the same rural home district in Chinhoyi, which is just north of Harare. Two of the patients infected with the Beijing/W strain also gave a history of a household contact with TB within the last 10 years, but none harbored a drug-resistant isolate.

DISCUSSION

We have used spoligotyping and VNTR typing of an epidemiologically well characterized patient population to describe the current pattern of TB transmission in Harare, Zimbabwe, a country in which HIV and TB are endemic. On the basis of shared spoligotype and VNTR patterns, more than three-quarters of the strains were potentially linked, suggesting a high level of recent TB transmission in Harare. This is substantially higher than the estimates of recent transmission reported from other studies. The proportions of new cases due to recent infection have been estimated in cities and countries where the rates of endemic TB are low and have been found to be 38 to

TABLE 4. Frequency of spoligotypes in Harare, Zimbabwe

Octal code[a]	No. of isolates observed in Harare	International Spoligotype database code	Octal code[a]	No. of isolates observed in Harare	International Spoligotype database code[b]
000000000003771	4	1	777737774020771	1	218
077777606060771	1	816	777737777760731	1	73
100775747413771	1	Not seen	777737777760771	4	37
457347607760671	1	Not seen	777747606060771	2	808
477777607760771	4	753	777757606060731	2	82
477777777760771	3	804	777757777760671	1	83
517347606060661	1	79	777773604060731	1	84
557163777760671	1	Not seen	777774606060731	1	Not seen
557757606060771	1	85	777775606060771	3	809
577347606060631	1	Not seen	777775770020771	1	Not seen
577717606060731	1	Not seen	777776777760771	1	119
577756777760601	1	Not seen	777777414020731	1	Not seen
577757604040731	1	Not seen	777777602060771	1	810
577757606060731	1	Not seen	777777604000171	1	Not seen
577757606060771	1	87	777777604060731	8	811
577777603760771	1	Not seen	777777606060571	3	812
601777606060731	1	Not seen	777777606060631	5	813
601777606060771	1	Not seen	777777606060631	1	Not seen
637775777763770	1	Not seen	777777606060671	4	814
677777607760771	2	20	777777606060731	9	815
700046677760671	1	Not seen	777777606060771	68	59
700076777760771	1	92	777777607760731	1	60
700775747473771	1	Not seen	777777607760771	23	42
700777747413771	2	129	777777674063771	1	Not seen
703377400001771	3	21	777777676060771	1	Not seen
730377600003771	1	Not seen	777777677760771	6	291
737417606060731	2	805	777777747473771	1	Not seen
737777607030771	1	Not seen	777777757760771	1	44
737777677670371	1	Not seen	777777757763771	1	Not seen
757757606060771	1	Not seen	777777773413731	1	Not seen
757777606060771	1	Not seen	777777774020731	2	62
757777777413731	2	806	777777774020771	1	47
770077606060731	3	807	777777774060771	1	Not seen
773777606060731	1	Not seen	777777777413771	1	236
774017606060731	1	Not seen	777777777420731	1	817
774777777760771	1	801	777777777473771	1	Not seen
777357606060731	1	Not seen	777777777600771	1	780
777357607760631	1	Not seen	777777777760671	1	245
777377607760771	1	81	777777777760771	25	53
777417606060731	9	184			

[a] Octal codes were determined by the protocol of Dale et al. (7a).
[b] Obtained from the Pasteur Institute, Pointe à Pitre, Guadeloupe.

41% in New York City (1, 14); 32% in Baltimore, Md. (6); 40% in San Francisco, Calif. (32); 28% in Berne, Switzerland (16); 46% in Amsterdam, The Netherlands (37); 28% in France (36); and 38% in Seville, Spain (31). Fewer molecular epidemiological studies of TB transmission have been undertaken in countries where HIV and TB are endemic; and estimates range from 25% of 84 TB patients in Honduras (30); 32% of 239 patients in San Paulo, Brazil, of whom approximately half were HIV infected (12); 33% of 51 patients in Guadeloupe (33); 41% in a study of 41 HIV-positive TB patients from Nairobi, Kenya (17); 42% of 301 acid-fast bacillus smear-positive patients from Botswana (69% were HIV seropositive) (27); 45% of 38 patients in Havana, Cuba (29); 45% of 246 patients (20) and 50% of 371 patients (42) in South Africa; 62% of patients in Tunisia (24); and 53.6% of 28 TB patients in a small study conducted in Harare in 1995 (25). Our high rate of clustering suggesting recent transmission probably reflects the high prevalence of HIV infection (74%) in our urban population and the relatively young age (median age, 31 years) of our urban population living in generally overcrowded, high-density housing. HIV infection increases susceptibility to exogenous TB infection (3), and in turn, primary TB is more likely to develop in the first 12 months following acquisition (3). Among HIV-infected patients living in New York City and San Francisco, more than 60% of new TB cases were the result of recent transmission (1, 32). These studies also found that the frequency of strains with clustered patterns was significantly higher among seropositive patients (1, 32), although this has not been a consistent finding (20, 27, 38).

The significance of clustering in population-based studies is controversial (18). While epidemiological data from urban areas strongly suggest that clustering indicates recent transmission of TB, in geographically stable rural populations, clustering may result from the simultaneous reactivation of infection acquired from the same source in the distant past (7) or from the predominance of certain well-conserved ancestral strains in the population, which would result in an overestimate of recent transmission. For example, in a study from Malawi and Kenya,

strains from a widespread area were found to have identical DNA fingerprints, but no apparent epidemiological links could be identified (21). We found that approximately half of our *M. tuberculosis* isolates that clustered by spoligotyping and that 70% of our *M. tuberculosis* isolates that clustered by analysis of VNTRs belonged to just a few major families of clustered patterns. A recent study of strains from countries with a high prevalence of *M. tuberculosis* infection has also suggested that the majority of circulating strains in a particular locality belong to a limited number of families (24), although Warren et al. (41) found a high degree of strain diversity in Cape Town, South Africa, where the incidence of TB is high. However, despite this, there was still substantial diversity in our typing patterns, with 54 different strains detected among our study population by spoligotyping and 33 different strains detected by analysis of VNTRs, suggesting that the chance occurrence of identical patterns among unrelated cases would be unusual. Furthermore, with 8 (35%) of 20 clusters identified by spoligotyping and 2 (17%) of the 12 clusters identified by analysis of VNTRs, there was a trend toward a higher rate of clustering within certain Harare districts, but statistical confirmation was limited by the small sample size. Although we did not undertake formal contact tracing to establish the presence of specific epidemiological links between these geographically based clusters, we found that almost all patients within these clusters lived within 0.5 km of each other. Importantly, we did not identify any clustering according to rural home district, which argues against the possibility that these clustered strains reflect reactivation of strains acquired from a common source in the past. Overall, these observations suggest that cases of TB caused by strains with identical spoligotype or VNTR patterns in our study are due to recently transmitted disease but that the incidence of TB in Harare is strongly influenced by a relatively small subset of actively circulating strains.

Various studies have identified multiple factors associated with TB due to recent infection, including young age (1, 30, 32); HIV infection status (1, 32); drug resistance (1, 12); and membership in ethnic minority populations, drug and alcohol use, and homelessness (1, 14, 16, 32, 37). The only risk factor for clustering in a recent molecular epidemiological study from Botswana comparable to our study was prior imprisonment (27). In our multivariate analysis, infection with an isolate with a pattern found to be part of a cluster was not associated with any particular demographic or clinical characteristics, including HIV infection. However, in the analysis of risk factors for membership in a large cluster (six or more strains), unique or nonclustered strains were significantly associated with a history of TB in the household in the last 5 years, and there was a borderline association with drug resistance. The lack of an association between HIV seropositivity and clustering has been reported in other studies (20, 27, 38, 42), and in our study it most likely reflects the very high HIV type 1 (HIV-1) seroprevalence in the population. Our explanation of the association between a household history of TB and infection with a nonclustered strain rather than a clustered strain is that residence within the various high-density districts of Harare is likely to be the more dominant factor in recent exposure and transmission and that reactivation of TB acquired from a household contact, often years previously, would result in a unique strain rather than a clustered strain. Finally, since we enrolled patients with new pulmonary cases of TB only over a 6-month period, it is unlikely that we would have been able to identify a link with a case from many years earlier.

Comparison with the International Spoligotype database demonstrated that our three largest spoligotype clusters are well recognized and ubiquitous in Africa. Type 59 is found almost exclusively in Africa and was described in a previous study in Harare (25). Types 53 and 42 are ubiquitous, with a trend for type 42 to be overrepresented in Latin America and Mediterranean countries, but probably also in Africa. We identified the presence of the Beijing/W genotype strain in four patients, none of whom were geographically or epidemiologically linked. Strains of the Beijing/W genotype family have had a strong impact on the TB epidemics in Asia, Vietnam, and the former republics of the USSR (40) and increasingly in other geographic regions (19). They have also been associated with large outbreaks, some involving multidrug resistance (19). The occurrence of the Beijing/W strain in Vietnam has been correlated with young age, suggesting recent transmission (2), and its presence in four patients in our study population is likely to represent imported transmission.

The spoligotyping and analysis of VNTR methods together confirmed the presence of a cluster of strains in 78.6% of patients, whereas 84.1% clustering was found on the basis of spoligotyping alone and 89.4% clustering was found on the basis of analysis of VNTRs alone. This high level of agreement between unlinked genetic markers is consistent with a clonal population structure of circulating *M. tuberculosis* strains (26). Analysis of VNTRs was less discriminatory as a first-line fingerprinting method: 11 of the 12 clusters identified by analysis of VNTRs could be further subdivided into distinct subtypes by spoligotyping, whereas only 8 of the 20 clusters identified by spoligotyping could be further subdivided by analysis of VNTRs. However, five spoligotypes did not have a corresponding VNTR fingerprint, so the discriminatory power of analysis of VNTRs may have been underestimated. Overall, the use of spoligotyping in tandem with analysis of VNTRs provided additional cluster discrimination and is comparable to that achieved by IS*6110* RFLP typing (13, 26). More recently, a promising new molecular typing method based on 12 minsatellite-like loci containing VNTRs of genetic elements, called mycobacterial interspersed repetitive units (MIRUs), has been developed (35). The resolution obtained by analysis of MIRUs-VNTRs is close to that of typing by RFLP analysis, and it is more discriminatory than spoligotyping or analysis of VNTRs alone (34). The level of clustering in our study may therefore have been less if the isolates had been typed by either the RFLP method or a two-PCR-based genotyping strategy of spoligotyping with MIRUs.

Our study has three main limitations affecting the generalizability of the data. First, we obtained DNA from only half of the patients enrolled in the study. However, the characteristics of the patients from whom isolates were and were not used for spoligotyping were similar, so this is unlikely to have biased our results. Similarly, although our study population was enrolled from only one hospital in Harare, the majority of TB patients in the city are seen at this hospital, and we therefore believe that the patients that we enrolled are representative of the TB patients in Harare. We also enrolled only smear-positive patients, and given that smear-negative but culture-positive cases

can account for considerable TB transmission, our study may provide an incomplete assessment of transmission within the community (5). However, by confining our analysis to smear-positive patients, we could identify the most efficient sources of TB transmission. Finally, we excluded laboratory cross-contamination as a likely reason for the clustering, as none of the samples with identical patterns were processed on the same day.

The study has several major implications for TB control. First, if chemoprophylaxis is to be used, it must be lifelong to cover the constant risk of exposure and not just short term to cover reactivation. Similarly, there may be little value in prolonging standard treatment to reduce the likelihood of relapse if the majority of cases are due to reinfection. Second, we assume that clustering occurs either because the contact is immunocompromised and progresses rapidly to disease after exposure or because the index case patient delays seeking medical attention, resulting in the potential for many secondary cases. If this is the case, then an increased emphasis should be placed on the early identification of cases through enhanced public health education campaigns, together with stricter implementation of the traditional methods of TB control, including treatment until cure.

ACKNOWLEDGMENTS

Financial support was provided by the Royal College of Physicians of London and a travel grant from the Wellcome Trust.

REFERENCES

1. **Alland, D., G. E. Kalkut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom.** 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N. Engl. J. Med. 330:1710–1716.
2. **Anh, D. D., M. W. Van Borgdorff, L. N. Van, N. T. Lan, T. van Gorkom, K. Kremer, and D. van Soolingen.** 2000. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. Emerg. Infect. Dis. 6:302–305.
3. **Barnes, P. F., A. B. Bloch, P. T. Davidson, and D. E. Snider, Jr.** 1991. Tuberculosis in patients with human immunodeficiency virus infection. N. Engl. J. Med. 324:1644–1650.
4. **Barnes, P. F., Z. Yang, S. Preston-Martin, J. M. Pogoda, B. E. Jones, M. Otaya, K. D. Eisenach, L. Knowles, S. Harvey, and M. D. Cave.** 1997. Patterns of tuberculosis transmission in central Los Angeles. JAMA 278: 1159–1163.
5. **Behr, M. A., S. A. Warren, H. Salamon, P. C. Hopewell, A. Ponce de Leon, C. L. Daley, and P. M. Small.** 1999. Transmission of *Mycobacterium tuberculosis* from AFB smear-negative patients. Lancet 353:444–449.
6. **Bishai, W. R., N. M. Graham, S. Harrington, D. S. Pope, N. Hooper, J. Astemborski, L. Sheely, D. Vlahov, G. E. Glass, and R. E. Chaisson.** 1998. Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA 280:1679–1684.
7. **Braden, C. R., G. L. Templeton, M. D. Cave, S. Valway, I. M. Onorato, K. G. Castro, D. Moers, Z. Yang, W. W. Stead, and J. H. Bates.** 1997. Interpretation of restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. J. Infect. Dis. 175:1446–1452.

7a. **Dale, J. W., D. Brittain, A. A. Cataldi, D. Cousins, J. T. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. A. Skuce, C. Sola, D. Van Soolingen, and V. Vincent.** 2001. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. Int. J. Tuberc. Lung Dis. 5:216–269.

8. **Daley, C. L., P. M. Small, G. F. Schecter, G. K. Schoolnik, R. A. McAdam, W. R. Jacobs, Jr., and P. C. Hopewell.** 1992. An outbreak of tuberculosis with accelerated progression among persons infected with human immunodeficiency virus: an analysis using restriction fragment length polymorphisms. N. Engl. J. Med. 326:231–235.
9. **DeCock, K. M., B. Soro, I. M. Coulibaly, and S. B. Lucas.** 1992. Tuberculosis and HIV infection in sub-Saharan Africa. JAMA 268:1581–1587.
10. **Dooley, S. W., M. E. Villarino, M. Lawrence, L. Salinas, S. Amil, J. V. Rullan, W. R. Jarvis, A. B. Bloch, and G. M. Cauthen.** 1992. Nosocomial transmission of tuberculosis in a hospital unit for HIV-infected patients. JAMA 267:2632–2634.
11. **Edlin, B. R., J. I. Tokars, M. H. Grieco, J. T. Crawford, J. Williams, E. M. Sordillo, K. R. Ong, J. O. Kilburn, S. W. Dooley, and K. G. Castro.** 1992. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N. Engl. J. Med. 326:1514–1521.
12. **Ferrazoli, L., M. Palaci, L. R. Marques, L. F. Jamal, J. B. Afiune, E. Chimara, M. C. Martins, M. A. Silva Telles, C. A. Oliveira, M. C. Palhares, D. T. Spada, and L. W. Riley.** 2000. Transmission of tuberculosis in an endemic urban setting in Brazil. Int. J. Tuberc. Lung Dis. 4:18–25.
13. **Filliol, I., S. Ferdinand, L. Negroni, C. Sola, and N. Rastogi.** 2000. Molecular typing of *Mycobacterium tuberculosis* based on variable number of tandem DNA repeats used alone and in association with spoligotyping. J. Clin. Microbiol. 38:2520–2524.
14. **Friedman, C. R., M. Y. Stoeckle, B. N. Kreiswirth, W. D. Johnson, Jr., S. M. Manoach, J. Berger, K. Sathianathan, A. Hafner, and L. W. Riley.** 1995. Transmission of multidrug-resistant tuberculosis in a large urban setting. Am. J. Respir. Crit. Care Med. 152:355–359.
15. **Frothingham, R., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem repeats. Microbiology 144:1189–1196.
16. **Genewein, A., A. Telenti, C. Bernasconi, C. Mordasini, S. Weiss, A. M. Maurer, H. L. Rieder, K. Schopfer, and T. Bodmer.** 1993. Molecular approach to identifying route of transmission of tuberculosis in the community. Lancet 342:841–844.
17. **Gilks, C. F., P. Godfrey-Faussett, B. I. Batchelor, J. C. Ojoo, S. J. Ojoo, R. J. Brindle, J. Paul, J. Kimari, M. C. Bruce, J. Bwayo, F. A. Plummer, and D. A. Warrell.** 1997. Recent transmission of tuberculosis in a cohort of HIV-1 infected female sex workers in Nairobi, Kenya. AIDS 11:911–918.
18. **Glynn, J. R., J. Bauer, A. S. de Boer, M. W. Borgdorff, P. E. Fine, P. Godfrey-Faussett, and E. Vynnycky.** 1999. Interpreting DNA fingerprint clusters of *Mycobacterium tuberculosis*. Int. J. Tuberc. Lung Dis. 3:1055–1060.
19. **Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen.** 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*. Emerg. Infect. Dis. 8:843–849.
20. **Godfrey-Faussett, P., P. Sonnenberg, S. C. Shearer, M. C. Bruce, C. Mee, L. Morris, and J. Murray.** 2000. Tuberculosis control and molecular epidemiology in a South African gold-mining community. Lancet 356:1066–1071.
21. **Godfrey-Faussett, P., and N. G. Stoker.** 1992. Aspects of tuberculosis in Africa. Genetic fingerprinting for clues to the pathogenesis of tuberculosis. Trans. R. Soc. Trop. Med. Hyg. 86:472–475.
22. **Goguet de la Semoniere, Y. O., H. M. Li, G. Torrea, A. Bunschoten, J. van Embden, and B. Gicquel.** 1997. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 35:2210–2214.
23. **Hayward, A. C., S. Goss, F. Drobniewski, N. Saunders, R. J. Shaw, M. Goyal, A. Swan, A. Uttley, A. Pozniak, J. Grace-Parker, and J. M. Watson.** 2002. The molecular epidemiology of tuberculosis in inner London. Epidemiol. Infect. 128:175–184.
24. **Hermans, P. W. M., F. Messadi, H. Guebrexabher, D. van Soolingen, P. E. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, and D. Frommel.** 1995. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J. Infect. Dis. 171:1504–1513.
25. **Heyderman, R. S., M. Goyal, P. Roberts, S. Ushewokunze, S. Zizhou, B. G. Marshall, R. Makombe, J. D. Van Embden, P. R. Mason, and R. J. Shaw.** 1998. Pulmonary tuberculosis in Harare, Zimbabwe: analysis by spoligotyping. Thorax 53:346–350.
26. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden.** 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 37:2607–2628.
27. **Lockman, S., J. D. Sheppard, C. Braden, M. J. Mwasekaga, C. L. Woodley, T. A. Kenyon, N. J. Binkin, M. Steinman, F. Montsho, M. Kesupile-Reed, C. Hirschfeldt, M. Notha, T. Moeti, and J. W. Tappero.** 2001. Molecular and conventional epidemiology of *Mycobacterium tuberculosis* in Botswana: a population-based prospective study of 301 pulmonary tuberculosis patients. J. Clin. Microbiol. 39:1042–1047.
28. **Ministry of Health and Child Welfare, Department of Epidemiology and Disease Control.** 1994. Tuberculosis annual report, Zimbabwe. Ministry of Health and Child Welfare, Harare, Zimbabwe.
29. **Montoro, E., J. Valdivia, and S. Cardoso Leao.** 1998. Molecular fingerprinting of *Mycobacterium tuberculosis* isolates obtained in Havana, Cuba, by IS*6110* restriction fragment length polymorphism analysis and by the double-repetitive element PCR method. J. Clin. Microbiol. 36:3099–3102.
30. **Pineda-Garcia, L. A., A. Ferreira, and S. E. Hoffner.** 1997. DNA fingerprinting of *Mycobacterium tuberculosis* strains from patients with pulmonary tuberculosis in Honduras. J. Clin. Microbiol. 35:2393–2397.
31. **Safi, H., J. Aznar, and J. C. Palomares.** 1997. Molecular epidemiology of *Mycobacterium tuberculosis* strains isolated during a 3-year period (1993 to 1995) in Seville, Spain. J. Clin. Microbiol. 35:2472–2476.
32. **Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston, G. F. Schecter, C. L. Daley, and G. K. Schoolnik.** 1994. The epidemiology of

tuberculosis in San Francisco: a population-based study using conventional and molecular methods. N. Engl. J. Med. 330:1703–1709.
33. **Sola, C. L., K. S. Horgen, S. Goh, and N. Rastogi.** 1997. Molecular fingerprinting of *Mycobacterium tuberculosis* on a Caribbean island with IS6110 and DRr probes. J. Clin. Microbiol. 35:843–846.
34. **Sola, C. L., I. Filliol, E. Legrand, S. Lesjean, C. Locht, P. Supply, and N. Rastogi.** 2002. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infect. Genet. Evol. 3:125–133.
35. **Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht.** 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J. Clin. Microbiol. 39:3563–3571.
36. **Torrea, G., C. Offredo, M. Simonet, B. Gicquel, P. Berche, and C. Pierre-Audigier.** 1996. Evaluation of tuberculosis transmission in a community by 1 year of systematic typing of *Mycobacterium tuberculosis* clinical isolates. J. Clin. Microbiol. 34:1043–1049.
37. **Van Deutekom, H., J. J. J. Gerritsen, D. van Soolingen, E. J. van Ameijden, J. D. van Embden, and R. A. Coutinho.** 1997. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. Clin. Infect. Dis. 25:1071–1077.
38. **Van Soolingen, D., M. W. Borgdorff, P. E. de Haas, M. M. Sebek, J. Veen, M. Dessens, K. Kremer, and J. D. van Embden.** 1999. Molecular epidemiology of tuberculosis in The Netherlands: a nationwide study from 1993 through 1997. J. Infect. Dis. 180:726–736.
39. **Van Soolingen, D., P. W. M. Hermans, P. E. W. de Haas, D. R. Soll, and J. D. van Embden.** 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J. Clin. Microbiol. 29:2578–2586.
40. **Van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden.** 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. 33:3234–3238.
41. **Warren, R., J. Hauman, N. Beyers, M. Richardson, H. S. Schaaf, P. Donald, and P. van Helden.** 1996. Unexpectedly high strain diversity of *Mycobacterium tuberculosis* in a high incidence community. S. Afr. Med. J. 86:45–49.
42. **Wilkinson, D., M. Pillay, J. Cramp, G. R. Davies, and A. W. Sturm.** 1997. Molecular epidemiology and transmission dynamics of *Mycobacterium tuberculosis* in rural Africa. Trop. Med. Int. Health 2:747–753.
43. **Wilson, S. A., S. Gross, and F. Drobniewski.** 1998. Evaluation of strategies for molecular fingerprinting for their use in the routine work of a *Mycobacterium* reference unit. J. Clin. Microbiol. 36:3385–3388.

JOURNAL OF CLINICAL MICROBIOLOGY, Jan. 2004, p. 431–434 Vol. 42, No. 1
0095-1137/04/$08.00+0 DOI: 10.1128/JCM.42.1.431–434.2004

Molecular Epidemiology of Disease Due to *Mycobacterium bovis* in Humans in the United Kingdom

Andrea L. Gibson,[1]* Glyn Hewinson,[2] Tony Goodchild,[2] Brian Watt,[3]† Alistair Story,[4] Jacqueline Inwald,[2] and Francis A. Drobniewski[1]

Mycobacterium Reference Unit, Health Protection Agency, and Department of Microbiology, Guy's, King's and St Thomas' School of Medicine, King's College Hospital (Dulwich), East Dulwich Grove, London, SE22 8QF,[1] Veterinary Laboratory Agency, Weybridge, New Haw, Surrey KT15 3NB,[2] Scottish Mycobacteria Reference Laboratory, Royal Infirmary, Little France, Edinburgh EH16 4SA,[3] and HPA, Communicable Disease Surveillance Centre, Colindale, London NW9 5EQ,[4] United Kingdom

Received 2 May 2003/Returned for modification 13 June 2003/Accepted 12 September 2003

***Mycobacterium bovis* is the causative agent of bovine tuberculosis, with a wide host range. Fifty human *M. bovis* isolates were typed using spoligotyping and variable number tandem repeats (VNTR). Fifteen of these spoligotypes have not yet been recorded in cattle. The predominant spoligotype in humans and cattle was subdivided by VNTR.**

Mycobacterium bovis has a wide host range, infecting many domestic and wild animals. Although occurring relatively rarely, *M. bovis* can also infect humans. In the United Kingdom, only about 1% of clinically diagnosed cases of tuberculosis (TB) that are subsequently proven bacteriologically are attributed to *M. bovis*, but in the developing world, *M. bovis* is still a cause for concern (6). The resurgence of bovine TB in cattle in the United Kingdom is raising concerns that transmission from cattle to humans might be a serious public health issue. It is therefore important to be able to quickly identify where rates of *M. bovis* in cattle are high and pose a potential risk of transmission to humans. *M. bovis* was once a major source of TB in humans in the United Kingdom but was almost eradicated after the introduction of control measures to reduce bovine tuberculosis in cattle together with the pasteurization of milk for human consumption. The majority of bovine TB cases in the 1980s and early 1990s presented either in the elderly or in those who had been infected abroad and returned or migrated to the United Kingdom (13). Many animals, such as badgers, foxes, ferrets, and deer (1, 3, 9), are believed to act as vectors for transmission to livestock, and some have also been associated with transmission to humans (8, 16, 18). Enhanced surveillance of *M. bovis* infections in humans was initiated in 1998. However, in 2001 a revised system which allows more timely collection of data was introduced (4, 5). Advances in molecular typing have provided tools to enhance our knowledge of *M. bovis* dissemination. Restriction fragment length polymorphism using the insertion sequence IS*6110* is considered to provide the best discrimination of *M. tuberculosis* isolates. However, *M. bovis* isolates from cattle usually have a single copy of IS*6110* (7); therefore, alternative techniques such as spacer oligonucleotide typing (spoligotyping) and variable number tandem repeats (VNTR) have been used successfully in discriminating between strains of *M. bovis* (1, 7, 11, 12, 15, 17).

This study examines the molecular epidemiology of *M. bovis* cases within the United Kingdom using two molecular typing techniques and compares the typing patterns obtained to those prevalent in United Kingdom cattle today.

All available viable *M. bovis* isolates (50 isolates) from humans diagnosed in the United Kingdom between 1997 and 2000 were identified; 40 were recovered at the Mycobacterium Reference Unit, London, and 10 were recovered at the Scottish Mycobacteria Reference Laboratory, Edinburgh. DNA was extracted by using a quick extraction method (19). Briefly, one colony was removed using a 1-μl loop and placed in 150 ml of water. An equal volume (150 ml) of chloroform was added, and the mixture was vortexed and then boiled at 80°C for 20 min to kill the cultures.

Spoligotyping was performed using the method described by Kamerbeek et al. (15), and VNTR was performed using the method described by Frothingham and Meeker-O'Connell (11). The size of each exact tandem repeat at each locus (A to E) was determined by running the PCR product on an agarose gel containing size markers (100-bp ladder; Promega, Southampton, United Kingdom) (20-bp ladder; Sigma-Aldrich, Dorset, United Kingdom). Deletion typing was carried out on a strain with a spoligotype not typical of *M. bovis*, using the method described by Brosch et al. (2). Seven regions of difference (RD) were examined: RD 4, 7, 8, 9, 10, 12, and 13. The Hunter-Gaston index (HGI), which is based on the probability of two unrelated strains from a test population being placed into different typing groups, was calculated to determine the discriminatory power of each typing method alone and in combination (14).

Epidemiological information was obtained from internal laboratory records at the Mycobacterium Reference Unit and Scottish Mycobacteria Reference Laboratory and from existing surveillance data held at the Health Protection Agency Communicable Disease Surveillance Centre.

Spoligotyping of the 50 human *M. bovis* isolates produced 25

* Corresponding author. Mailing address: Mycobacterium Reference Unit, Health Protection Agency, Dulwich Hospital, East Dulwich Grove, London SE22 8QF, United Kingdom. Phone: 0208 693 1312. Fax: 0207 346 6477. E-mail: andreagibson78@hotmail.com.

† Present address: Silverburn House, by Penicuik, Midlothian, EH26 9LF, United Kingdom.

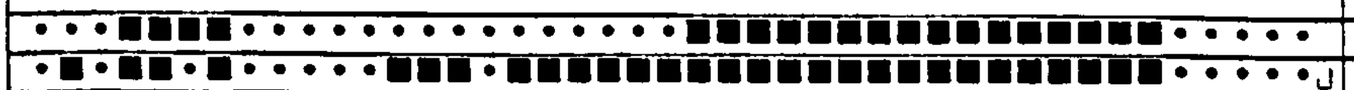
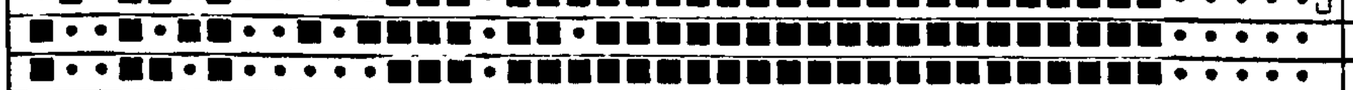
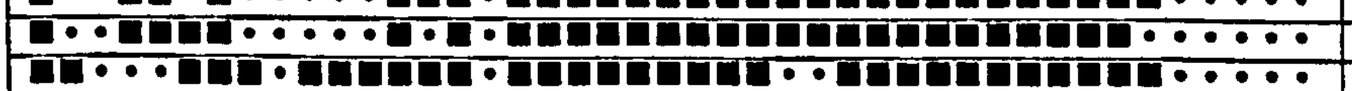
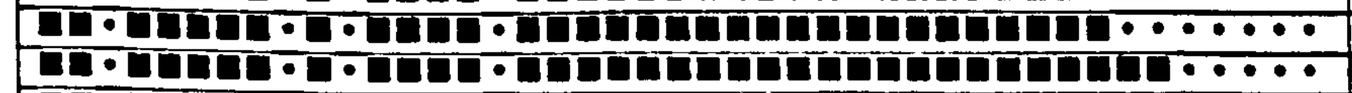
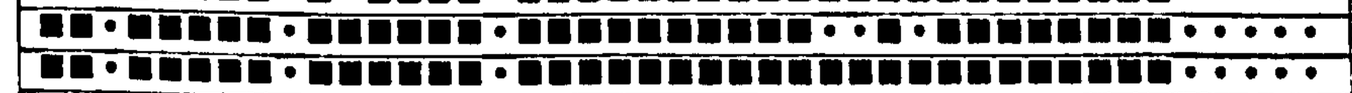
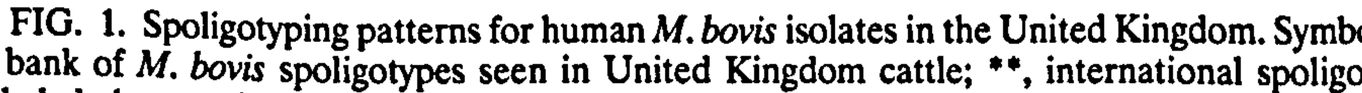
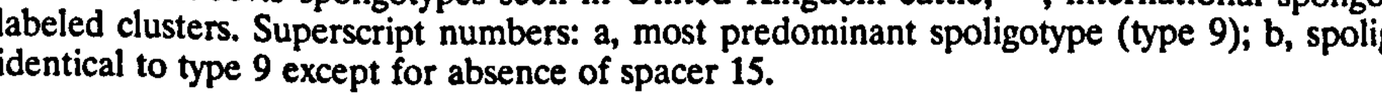
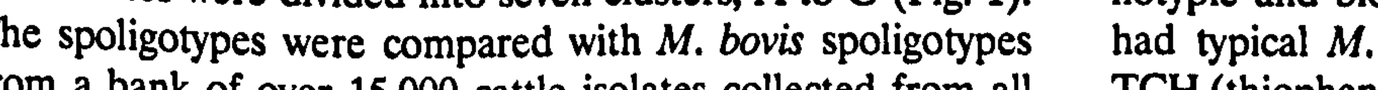
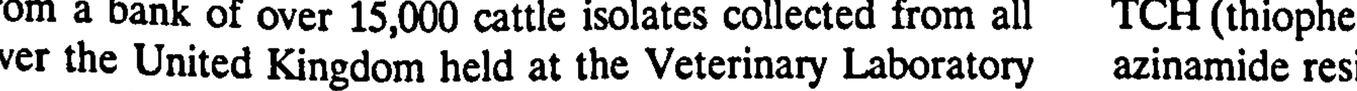
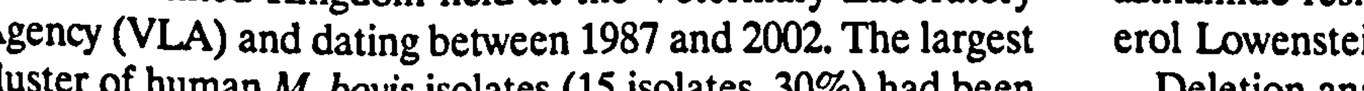
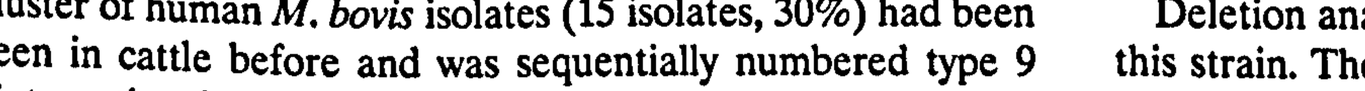
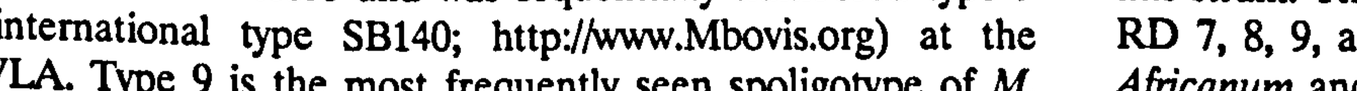
Human *M. bovis* spoligotypes	VLA* Type	Internati-onal Type**	No. of isolates	Cluster†
	31	SB0155	1	
	22	SB0673	1	
	Ns[b]	Ns	1	
	Ns	Ns	1	
	Ns	Ns	1	
	Ns	Ns	1	
	35	SB0134	1	
	Ns	Ns	1	
	20	SB0145	1	
	6	SB0267	1	
	Ns	SB0131	2	A
	9[c]	SB0269	2	B
	Ns	SB0144	1	
	Ns	Ns	1	
	Ns	Ns	2	C
	17	SB0263	1	
	10	SB0272	3	D
	13	SB0273	1	
	9[a]	SB0140	15	E
	Ns	Ns	1	
	Ns	Ns	1	
	21	SB0130	6	F
	Ns	Ns	1	
	BCG	SB0120	2	G
	Ns	Ns	1	

FIG. 1. Spoligotyping patterns for human *M. bovis* isolates in the United Kingdom. Symbols: *, comparison of human *M. bovis* spoligotypes with a bank of *M. bovis* spoligotypes seen in United Kingdom cattle; **, international spoligotype website is http://www.Mbovis.org; –T arbitrarily labeled clusters. Superscript numbers: a, most predominant spoligotype (type 9); b, spoligotype not seen in United Kingdom cattle before; c, identical to type 9 except for absence of spacer 15.

individual spoligotypes (Fig. 1) and had an HGI of 0.90. Thirty-two isolates were divided into seven clusters, A to G (Fig. 1). The spoligotypes were compared with *M. bovis* spoligotypes from a bank of over 15,000 cattle isolates collected from all over the United Kingdom held at the Veterinary Laboratory Agency (VLA) and dating between 1987 and 2002. The largest cluster of human *M. bovis* isolates (15 isolates, 30%) had been seen in cattle before and was sequentially numbered type 9 (international type SB140; http://www.Mbovis.org) at the VLA. Type 9 is the most frequently seen spoligotype of *M. bovis* (over 30% of all isolates have this spoligotype) isolated from cattle and has a wide geographical range in the United Kingdom (10).

Human type 9 isolates were seen across the United Kingdom, suggesting that transmission between cattle and humans might occur. Interestingly, 15 of the human *M. bovis* spoligotypes had not been seen at the VLA in isolates from cattle. When these 15 types were compared to the international spoligotype database, only 2 were recognized. The first was isolated in Argentina, the second was isolated in Australia, and the remaining 13 spoligotypes were all unique to the United Kingdom. In the majority of these cases, it is likely that disease was due to reactivation of a past infection that had been acquired prior to milk pasteurization rather than to primary infection, because 72.3% of the patients were over the age of 50. (Fig. 2). Therefore, these 13 unique spoligotypes may reflect *M. bovis* strains circulating in the United Kingdom over 50 years ago.

M. bovis spoligotypes do not usually contain spacers 39 to 43; however, one spoligotype from the panel contained spacers 40 to 43, which are more commonly seen in *M. tuberculosis*. Phenotypic and biochemical tests demonstrated that this isolate had typical *M. bovis* characteristics; it was microaerophilic, TCH (thiophen-2-carboxylic acid hydrazide) negative, and pyrazinamide resistant, and it grew better on pyruvate than glycerol Lowenstein-Jensen slopes.

Deletion analysis was performed to ascertain the identity of this strain. The strain contained RD 4, 12, and 13 but lacked RD 7, 8, 9, and 10, indicating that this strain is actually *M. Africanum* and not *M. bovis*.

VNTR typing alone produced 18 different patterns and had an HGI of 0.85. Combining spoligotyping with VNTR vastly improved the level of discrimination, producing 34 different types and a very high HGI of 0.96. Furthermore, VNTR was very useful in subdividing type 9 spoligotypes, separating the group into six subtypes (Table 1).

Epidemiological information showed that the study population was widely distributed across the United Kingdom, had an average age of 58.7 years, and had approximately equal proportions of males and females (21:17). Where ethnicity was known (a total of 15 patients), 14 patients were white and 1 was of black-African origin and was originally from Nigeria but had lived in the United Kingdom since 1996. This person had a unique spoligotype; therefore, it is possible that she was infected in Nigeria before arriving in the United Kingdom. Of interest, 59% (13 of 22) of cases had some contact with a farm, ranging from having a Saturday job milking cows, to living on a dairy farm as a child, to being a farmer (now retired). One spoligotype cluster represents an outbreak on a farm in Gloucester. Two siblings (a 20-year-old male and a 17-year-old

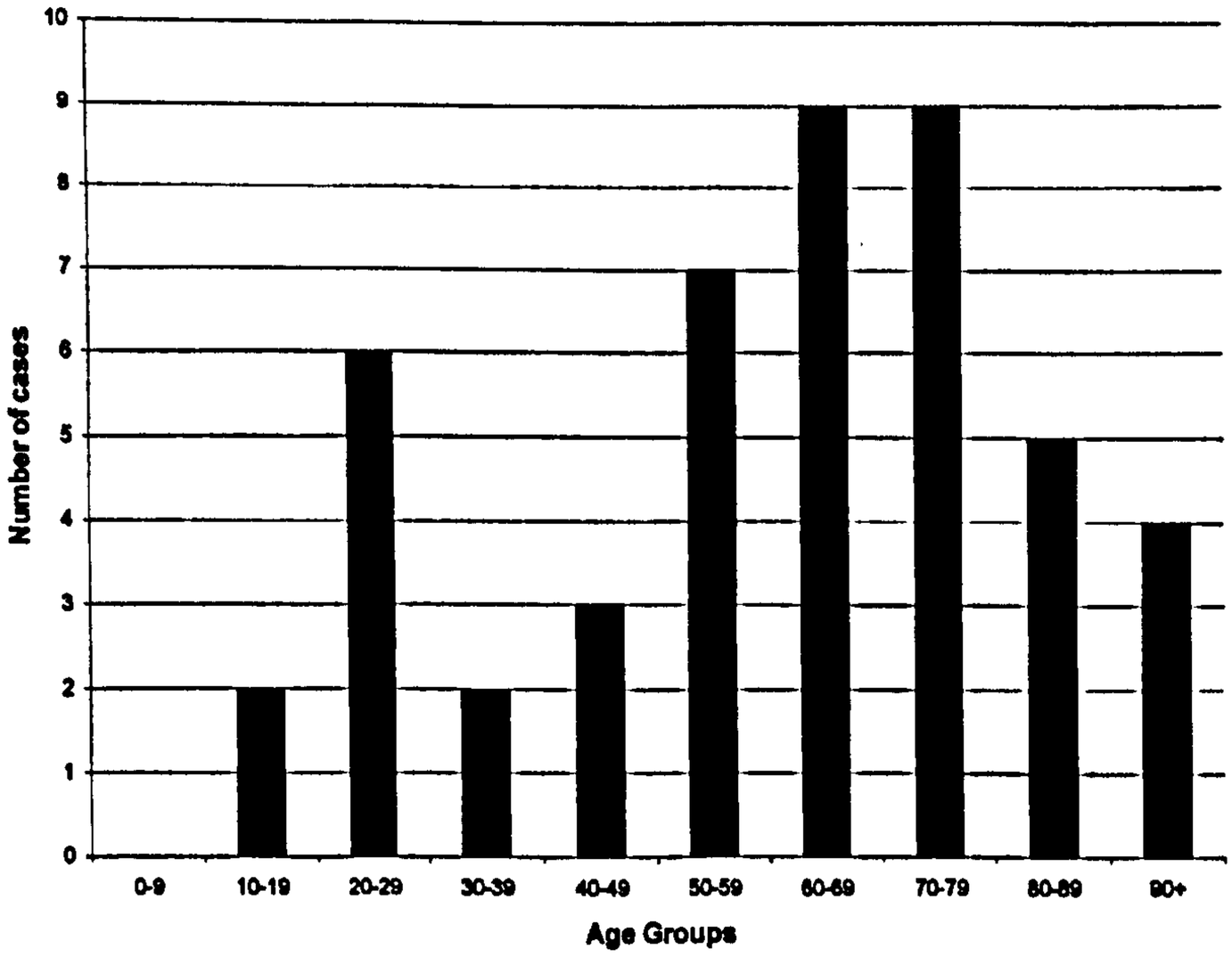


FIG. 2. Number of *M. bovis* cases by age group.

female) living on their parents' farm became infected with *M. bovis*. The brother occasionally helped his father on the farm by restraining the cattle and would often be sprayed with nasal mucus. Cattle infected with *M. bovis* of the same spoligotype had been detected on the farm in previous years. Transmission from cattle to human is thought to have occurred by the inhalation of infected aerosols from cattle. The brother is thought to have subsequently infected his sister, as she had no contact with the cattle but was also diabetic and pregnant, i.e., immuno-compromised. This is thought to be the first case of human-to-human transmission since 1990 (R. M. M. Smith, F. Drobniewsky, A. L. Gibson, J. D. E. Montague, M. N. Logan, D. Hunt, R. G. Hewinson, R. L. Salmon, and B. O'Neill, unpublished data).

It is important to monitor bovine tuberculosis in humans, especially in those who are at high risk of primary infection, such as agricultural and abattoir workers, and to identify any transmission between animals and humans. A combination of spoligotyping and VNTR is an efficient discriminatory tool for the molecular surveillance of *M. bovis* and also addresses the problem of analyzing isolates with single copies of IS*6110*. The combined VNTR and spoligotyping approach is of value in typing *M. tuberculosis* isolates. Further improvements in these techniques might produce a combined system capable of high discrimination for all *M. tuberculosis* complex isolates in humans or other mammals.

TABLE 1. Subdivision of spoligotyping clusters by VNTR

Spoligo cluster[a] (no. of isolates)	No. of VNTR subtypes	VNTR profiles
A (2)	2	55543 75543
B (2)	2	55543 75543
C (2)	1	75553
D (3)	2	63543 75543
E (15)	6	56543 63543 65542 65543 75543 75553
F (6)	2	65543 66543
G (2)	2	54544 55343

[a] Arbitarily labeled clusters.

This work was funded by the Department for Environment, Food, and Rural Affairs and supported by the Health Protection Agency, formerly the Public Health Laboratory.

REFERENCES

1. **Aranaz, A., E. Liebana, A. Mateos, L. Dominguez, D. Vidal, M. Domingo, O. Gonzolez, E. F. Rodriguez-Ferri, A. E. Bunschoten, J. D. van Embden, and D. Cousins.** 1996. Spacer oligonucleotide typing of *Mycobacterium bovis* strains from cattle and other animals: a tool for studying epidemiology of tuberculosis. J. Clin. Microbiol. 34:2734–2740.
2. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc. Natl. Acad. Sci. USA 99:3684–3689.
3. **Clifton-Hadley, R. S., J. W. Wilesmith, M. S. Richards, P. Upton, and S. Johnston.** 1995. The occurrence of Mycobacterium bovis infection in cattle in and around an area subject to extensive badger (Meles meles) control. Epidemiol. Infect. 114:179–193.
4. **Communicable Disease Report Weekly.** 1998. Enhanced surveillance of *Mycobacterium bovis* in humans. Commun. Dis. Rep. Wkly. 8:281, 284.
5. **Communicable Disease Report Weekly.** 2001. Enhanced surveillance of *Mycobacterium bovis* disease in humans in England and Wales from January 2001. Commun. Dis. Rep. Wkly. 11:2–3.
6. **Cosivi, O., J. M. Grange, C. J. Daborn, M. C. Raviglione, T. Fujikura, D. Cousins, R. A. Robinson, H. F. A. K. Huchzermeyer, I. de Kantor, and F. X. Meslin.** 1998. Zoonotic tuberculosis due to Mycobacterium bovis in developing countries. Emerg. Infect. Dis. 4:59–70.
7. **Cousins, D. V., R. A. Skuce, R. R. Kazwala, and J. D. van Embden.** 1998. Towards a standardized approach to DNA fingerprinting of Mycobacterium bovis. International Union Against Tuberculosis and Lung Disease, Tuberculosis in Animals Subsection. Int. J. Tuberc. Lung Dis. 2:471–478.
8. **Dalovisio, J. R., M. Setter, and S. Mikota-Wells.** 1992. Rhinoceros' rhinorrhea: cause of an outbreak of infection due to airborne Mycobacterium bovis in zookeepers. Clin. Infect. Dis. 15:598–600.
9. **Delahay, R. J., C. L. Cheeseman, and R. S. Clifton-Hadley.** 2001. Wildlife disease reservoirs: the epidemiology of Mycobacterium bovis infection in the

European badger (Meles meles) and other British mammals. Tuberculosis 81:43–49.
10. **Durr, P. A., R. S. Clifton-Hadley, and R. G. Hewinson.** 2000. Molecular epidemiology of bovine tuberculosis. II. Applications of genotyping. Rev. Sci. Tech. 19:689–701.
11. **Frothingham, F., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem repeats. Microbiology 144:1189–1196.
12. **Haddad, N., A. Ostyn, C. Karoui, M. Masselot, M. F. Thorel, S. L. Hughes, J. Inwald, R. G. Hewinson, and B. Durand.** 2001. Spoligotype diversity of *Mycobacterium bovis* strains isolated in France from 1979 to 2000. J. Clin. Microbiol. 39:3623–3632.
13. **Hardie, R. M., and J. M. Watson.** 1992. Mycobacterium bovis in England and Wales: past, present and future. Epidemiol. Infect. 109:23–33.
14. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26:2465–2466.
15. **Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. 35:907–914.
16. **Michalak, K., C. Austin, S. Diesel, M. J. Bacon, P. Zimmerman, and J. N. Maslow.** 1998. Mycobacterium tuberculosis infection as a zoonotic disease: transmission between humans and elephants. Emerg. Infect. Dis. 4:283–287.
17. **Roring, S., A. Scott, D. Brittain, I. Walker, G. Hewinson, S. Neill, and R. Skuce.** 2002. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J. Clin. Microbiol. 40:2126–2133.
18. **van Soolingen, D., P. E. de Haas, J. Haagsma, T. Eger, P. W. Hermans, V. Ritacco, A. Alito, and J. D. van Embden.** 1994. Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. J. Clin. Microbiol. 32:2425–2433.
19. **Wilson, S. M., R. McNerney, P. M. Nye, P. D. Godfrey-Faussett, N. G. Stoker, and A. Voller.** 1993. Progress toward a simplified polymerase chain reaction and its application to diagnosis of tuberculosis. J. Clin. Microbiol. 31:776–782.

Mycobacterium bovis Infection, United Kingdom

Robert M.M. Smith,* Francis Drobniewski,† Andrea Gibson,† John D.E. Montague,‡ Margaret N. Logan,§ David Hunt,¶ Glyn Hewinson,# Roland L. Salmon,* and Brian O'Neill**

We describe the first documented spillover of bovine tuberculosis from animals into the human population of the United Kingdom since the resurgence of the disease in cattle in the country. This finding suggests that there may be a small risk for transmission to humans, making continued vigilance particularly necessary.

In the past, *Mycobacterium bovis* was a major source of tuberculosis in humans through consumption of unpasteurized milk. Currently, tuberculosis as a result of *M. bovis* infection is comparatively rare, but it remains a cause for concern in persons at high risk, such as abattoir workers (1). *M. bovis* principally affects cattle, but it can cause disease in a range of wild and domesticated animals, for example, badgers, ferrets, cats, deer, and llamoids (2). In U.K. cattle, *M. bovis* infection is now primarily a pulmonary disease, and the main route of transmission is likely to be through aerosol dissemination. Currently, approximately 1% of human tuberculosis cases can be attributed to *M. bovis*; most of those are likely to follow reactivation (3) or to be recent infections contracted abroad. Tuberculosis caused by *M. bovis* in the young is usually a primary infection. We report two human cases in Gloucestershire. One of the cases may have resulted from intrafamilial spread.

Case Studies

Bovine tuberculosis was diagnosed in two siblings—one currently residing, the other residing until recently, on their parents' farm in Gloucestershire— in 1999. A brother and sister ages 20 and 17 years, respectively, they are thought to have the first cases of indigenously acquired bovine tuberculosis caused by *M. bovis* in persons <25 years of age, with no documented history of travel abroad, reported to the Public Health Laboratory Service Communicable Disease Surveillance Centre (CDSC) since 1990.

When seen by her physician in 1999, the 17-year-old girl had a 6-month history of cough, weight loss, and lethargy. Infection with acid- and alcohol-fast bacilli (AAFB) was confirmed by culture of bronchial washings. Her brother had an 18-month history of cough. He was subsequently found to be AAFB-smear positive with pulmonary cavitation (i.e., he had an infectious case). Sputum samples from both case-patients were cultured by the Gloucester Public Health Laboratory, and the cultures were identified as *M. bovis* by the Regional Centre for Mycobacteriology in Cardiff. All human *M. tuberculosis* isolates are subjected to biochemical analysis and pyrazinamide drug susceptibility testing for differentiation of *M. bovis*. *M. bovis* is therefore detected as part of the routine reference service.

Both siblings had lived on the same farm most of their lives. However, the sister had recently moved into her own place at the time of her diagnosis. Both patients had received *Mycobacterium bovis* BCG in secondary school. Both smoked. Neither had knowingly drunk unpasteurized milk. The girl had no cattle contact. Her brother had occasional cattle exposure: he would assist when stock were confined in a cattle crush[1] for veterinary examination and restrained them by holding their nostrils. During this process, he could become covered in bovine mucus and saliva. He also reported contact with feral ferrets.

No disease has been reported in other family members or in social contacts. Results of screening of other family members (mother, father, and another sibling) were unremarkable. Their father had a grade 2 Heaf test result (and a previous history of BCG) (this is equivalent to a Mantoux response of induration of diameter 5 to 14 mm). Their mother had two grade 1 Heaf test results (Mantoux response of 0 to 4-mm induration) and no history of BCG. The other sibling (age 8) had a grade 1 and a grade 2 Heaf test result and no history of BCG. Heaf grades 0 and 1 or a Mantoux response of 0 to 4 mm induration are regarded as negative; those with a grade 2 reaction (or a Mantoux response of induration of diameter 5 to 14 mm following injection of 0.1 mL purified protein derivative 100 U/mL) are positive. Persons with a grade 2 response are hypersensitive to tuberculin protein and are not given BCG vaccination. A strongly positive reaction to tuberculin is

*National Public Health Service for Wales, Cardiff, United Kingdom; †Health Protection Agency, London, United Kingdom; ‡Department for Environment, Food and Rural Affairs, Gloucestershire, United Kingdom; §Gloucester Hospitals National Health Service Trust (formerly Gloucester Public Health Laboratory), Gloucester, United Kingdom; ¶Health Protection Agency (West Midlands) Regional Surveillance Unit, Birmingham, United Kingdom; #Veterinary Laboratories Agency, Weybridge, United Kingdom; and **Gloucestershire Health Protection Unit, Gloucester, United Kingdom

[1]A cattle crush is a small, often portable, cage used to restrain individual animals. Crushes are often used in the United Kingdom to hold animals for veterinary attention.

demonstrated by a Heaf grade 3 or 4 or a Mantoux response with induration of at least 15-mm diameter.

The farm had previously held a maximum of 25 beef cattle, introduced around 1981. Tuberculosis (*M. bovis*) herd breakdowns[2] had been recorded by the former Ministry of Agriculture, Fisheries and Food. Five cattle (of 7 slaughtered in a herd of 15) had bovine tuberculosis in 1993; they had caseous lymph node lesions and were culture positive. Another three (out of the herd of eight, which were all slaughtered) had similar lesions in 1997. All infected cattle showed lesions typical of *M. bovis* with confirmatory culture obtained; one had prescapular lymph nodes enlarged with caseous changes. The remaining animals were slaughtered as "direct contacts." Tuberculosis breakdowns have been reported in neighboring herds, and the area supports a substantial badger population. After the 1993 cattle breakdowns, five badgers were trapped; four were positive for *M. bovis* on culture. Similarly, in 1997, a single trapped badger was culture positive.

Conclusions

M. bovis from the cases and from cattle on the farm in 1997 were indistinguishable by a combination of restriction fragment length polymorphism (RFLP) analysis using the IS6110 element, spacer oligonucleotide ("spoligotyping"), and variable number tandem repeat (VNTR) analysis (4–6). RFLP analysis using the IS6110 insertion sequence represents the standard criterion for differentiating *M. tuberculosis,* but it is insufficiently discriminating for *M bovis* due to the paucity of IS6110 elements in the genome of this bacterium. Spoligotyping is based on the polymerase chain reaction (PCR) amplification of a polymorphic direct repeat (DR) locus in which the DR elements are interspersed with up to 43 spacer regions (Figure). The typing process relies on the presence or absence of spacers in the amplified DNA, which are detected by hybridization to a series of synthetic spacer oligonucleotides covalently linked to a filter. The presence of hybridized areas is shown by using a chemiluminescent reaction detected on film as a dark band; absence of spacers shows no binding. The sequence is then displayed as a binary bar code, which can be manipulated digitally. Similarly, VNTR analysis uses PCR to amplify a region in which there are tandem repeats at multiple loci. The result is a digital code describing the number of repeat units at each locus (Figure). The spoligotype profile obtained in these cases is one of the most common seen in bovine tuberculosis in the United Kingdom, and caution is needed before one can say unambiguously that strains have been transmitted. Nevertheless, the combination of typing methods, together with supportive epidemiology, provides evidence of exposure to a common source of infection.

M. bovis was characteristically transmitted to humans by ingestion of infected milk. Thus, historically, human *M. bovis* lesions were primarily extrapulmonary or intestinal. Cattle infected with *M. bovis*, by contrast, usually have pulmonary infection, and shedding of *M. bovis* in respiratory secretions has been reported by several workers (7–9). It is suggested that a possible route of badger to cattle transmission is by inhalation of bacilli from grass contaminated with infected badger urine, feces, or sputum (10). Cattle preferentially graze edges of fields, and they may sometimes be forced to graze close to badger latrines and scent-marking areas at the edge of fields. Cattle-to-cattle transmission of *M. bovis* is also likely to be important. Work to date (11) indicates that particular tuberculosis spoligotypes are usually clustered in specific areas, implying that herd breakdowns are localized events originating from a relatively static reservoir. In many instances, cattle and badgers have been found to share similar spoligotypes (11), but further sampling of badgers, cattle, and other wildlife is required to identify which species can share the infection. Current Department for Environment Food and Rural Affairs research is aimed at establishing the epidemiology and pathogenesis of *M. bovis* and the possible pathways of interspecies transmission.

Agricultural workers may acquire the disease by inhaling cough spray from infected cattle. Typical pulmonary tuberculosis then develops, which is what we believe occurred here.

Despite a long history of cattle herd breakdowns on this farm, the family members were not screened until the human cases occurred. Early detection of the disease in the young man before it became infectious might have prevented transmission to his sister and avoided the need for chemoprophylaxis for her infant son. No guidelines were in force at the time. Those subsequently issued (12) advocate screening of human contacts of disease only where pulmonary or udder lesions are detected in cattle. Since the early 1980s, reports of cattle herd breakdowns have steadi-

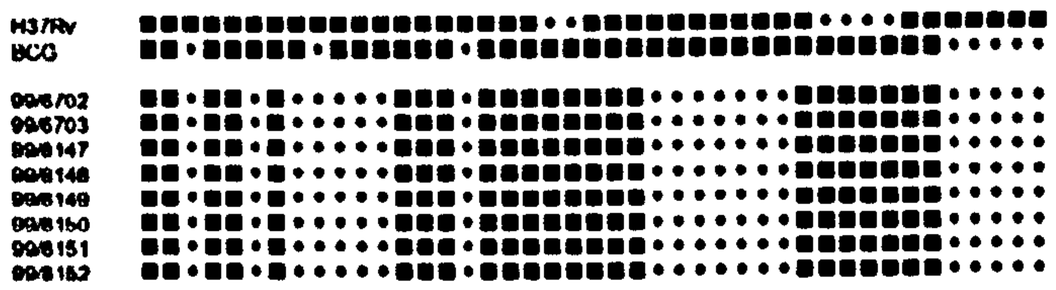


Figure. Spoligotyping profiles for human and cattle cases of bovine tuberculosis. H37Rv and BCG are control strains. 99/6702 and 99/6703 are from the sister and brother, respectively; 99/8147–99/8152 are cattle isolates.

[2]A tuberculosis breakdown is confirmed when a visible lesion, typical of tuberculosis, is seen by an official veterinary officer in the carcass of a tuberculosis reactor at postmortem or, if *Mycobacterium bovis* is cultured from the set of lymph glands that are routinely collected at postmortem and sent to a state veterinary laboratory for culture.

ly risen, with a more dramatic increase since 1990. The Southwest of England, the West Midlands, and South and West Wales have had recent increases in disease in cattle, and this trend is extending northward to include Derbyshire, Staffordshire, and Shropshire. This incident represents the first documented probable spillover into the human population from animals since the disease's resurgence in cattle, and it suggests there may be a small risk for transmission to humans, even when the bovine case is reported as closed,[3] because of the presence of *M. bovis* in the cattle's respiratory tract (7).

Acknowledgments

We thank Jeff Meecham-Jones, Malcolm Yates for restriction fragment length polymorphism typing, David Brown, and John Watson.

Dr. Smith is lead clinical scientist in the National Public Health Service for Wales Zoonoses Surveillance Unit, based in Cardiff. His research interests cover a wide range of indigenous and imported, zoonotic, and parasitic infections in the United Kingdom.

References

1. Grange JM, Yates MD. Zoonotic aspects of *Mycobacterium bovis* infection. Vet Microbiol 1994;40:137–51.
2. Krebs JR, Anderson R, Clutton-Brock R, Morrison I, Young D, Donnelly C. Bovine tuberculosis in cattle and badgers. (Krebs report). London: MAFF Publications; 1997.
3. Hardie RM, Watson JM. *Mycobacterium bovis* in England and Wales: past, present and future. Epidemiol Infect 1992;109:23–33.
4. Goyal M, Saunders NA, Van Embden JDA, Young DB, Shaw RJ. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS6110 restriction fragment length polymorphism. J Clin Microbiol 1997;35:647–51.
5. Kremer K, Van Soolingen D, Frothingham R, Haas WH, Hermans PWM, Martin C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
6. Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 1998;144:1189–96.
7. Neill SD, Hanna J, Mackie DP, Bryson TGD. Isolation of *Mycobacterium bovis* from the respiratory tracts of skin test-negative cattle. Vet Rec 1992;131:45–7.
8. Neill SD, O'Brien JJ, McCracken RM. *Mycobacterium bovis* in the anterior respiratory tracts in the heads of tuberculin-reacting cattle. Vet Rec 1988;122:184–6.
9. Cutbill LJ, Lynn A. Pulmonary tuberculosis of bovine origin. Br Med J 1944;i:283–9.
10. Hutchings MR, Harris S. Effects of farm management practices on cattle grazing. Behaviour and the potential for transmission of bovine tuberculosis from badgers to cattle. Vet J 1997;153:149–62.
11. Durr PA, Clifton-Hadley RS, Hewinson RG. Molecular epidemiology of bovine tuberculosis. II. Applications of genotyping. Rev Sci Tech 2000;19:689–701.
12. United Kingdom Departments of Health. Bovine tuberculosis: guidance on management of the public health consequences of tuberculosis in cattle in England and Wales. London: Department of Health; 2000.

[3]"Closed" refers to a gastrointestinal tract case in which the lymph nodes had not caseated and there was no perceived contact with the outside of the body cavity, as there would be a with a known respiratory (open) infection involving the lungs.

Address for correspondence: Robert Smith, Communicable Disease Surveillance Centre, National Public Health Service for Wales, Abton House, Wedal Road, Cardiff CF14 3QX, United Kingdom; fax: +44(0)29 20 521987; email: robert.smith@nphs.wales.nhs.uk

rich4/zam-aem/zam-aem/zam01205/zam6156d05z | xppws | S=1 | 10/19/05 | 10:40 | Art: 1165-05

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 2005, p. 001 Vol. 71, No. 12
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.12.000–000.2005

Can 15-Locus Mycobacterial Interspersed Repetitive Unit–Variable-Number Tandem Repeat Analysis Provide Insight into the Evolution of *Mycobacterium tuberculosis*?

Andrea Gibson, Timothy Brown, Lucy Baker,† and Francis Drobniewski*

Mycobacterium Reference Unit, Health Protection Agency, Barts and The London Queen Mary's School of Medicine and Dentistry, London, United Kingdom

Received 19 May 2005/Accepted 12 August 2005

The phylogeny and evolution of the bacterium *Mycobacterium tuberculosis* is still poorly understood despite the application of a variety of molecular techniques. We analyzed 469 *M. tuberculosis* and 49 *Mycobacterium bovis* isolates to evaluate if the mycobacterial interspersed repetitive units–variable-number tandem repeats (MIRU-VNTR) commonly used for epidemiological studies can define the phylogeny of the *M. tuberculosis* complex. This population was characterized by previously identified silent single-nucleotide polymorphisms (sSNPs) or by a macroarray based on these sSNPs that was developed in this study. MIRU-VNTR phylogenetic codes capable of differentiating between phylogenetic lineages were identified. Overall, there was 90.9% concordance between the lineages of isolates as defined by the MIRU-VNTR and sSNP analyses. The MIRU-VNTR phylogenetic code was unique to *M. bovis* and was not observed in any *M. tuberculosis* isolates. The codes were able to differentiate between different *M. tuberculosis* strain families such as Beijing, Delhi, and East African-Indian. Discrepant isolates with similar but not exactly identical MIRU-VNTR codes often displayed a stepwise trend suggestive of bidirectional evolution. A lineage-specific panel of MIRU-VNTR can be used to subdivide each lineage for epidemiological purposes. MIRU-VNTR is a valuable tool for phylogenetic studies and could define an evolutionarily uncharacterized population of *M. tuberculosis* complex organisms.

Despite the application of several molecular markers, relatively little is known about the evolution of *Mycobacterium tuberculosis* (3, 5, 10, 27), the bacterial species which causes 2 million deaths and 8 million new cases annually (8, 25).

Genotyping techniques based on neutral genetic variation such as multilocus sequence typing (MLST) have been used successfully to characterize bacterial populations (6, 9, 15, 20). Although the *M. tuberculosis* genome is thought to be highly conserved (23, 28), sufficient neutral variation was found within genes associated with drug resistance in *M. tuberculosis* complex isolates for construction of a robust phylogenetic tree, demonstrating that *M. tuberculosis* is clonal and is subdivided into four distinct lineages while *Mycobacterium bovis* is closely related but is found on a separate branch (1).

MIRU-VNTR (mycobacterial interspersed repetitive units–variable-number tandem repeats) is a high-throughput technique that analyzes the number of tandem repeats at loci distributed around the *M. tuberculosis* genome (30). Studies have explored MIRU-VNTR as a tool for discriminating between strains and in comparison with other molecular techniques (4, 16, 19, 21, 26, 30), but there are few published reports on the use of MIRU-VNTR to study the evolution of the *M. tuberculosis* complex. VNTR analysis using 5-locus and 12-locus MIRU has been used in combination with other techniques such as spoligotyping to investigate the evolution of *M. tuberculosis* complex (11, 17, 27), but to our knowledge, no studies have utilized 15-locus MIRU-VNTR to define evolutionary pathways characterized by silent single-nucleotide polymorphism (sSNP) analysis.

Based on the SNP-defined phylogenetic tree (1), we evaluate the use of MIRU-VNTR, commonly used in epidemiological investigations, as a rapid tool for the phylogenetic classification of *M. tuberculosis* and *M. bovis*. To validate the MIRU-VNTR phylogenetic classification, an additional panel of *M. tuberculosis* and *M. bovis* strains was characterized with an sSNP macroarray tool (the development of which is described here) and concordance with MIRU-VNTR was examined.

MATERIALS AND METHODS

Mycobacterial strains. A panel (panel 1, test population) of 312 *M. tuberculosis* clinical isolates collected in England and Wales between 1 January and 31 December 1998, which included all isolates resistant to one or both of the first-line antituberculous drugs (isoniazid and rifampin) and 100 randomly chosen fully susceptible isolates, as well as 4 *M. bovis* isolates (1), was typed using MIRU-VNTR (12 MIRU plus exact tandem repeats [ETR] A, B, and C; the two remaining ETR, D and E, are included within the 12 MIRU and correspond to MIRU 4 and 31). All members of this panel had previously been classified into one of five lineages according to the presence of four characteristic sSNPs (1). The laboratory reference strain *M. tuberculosis* H37Rv (which had been characterized by sSNP) was also MIRU-VNTR typed.

A second panel (panel 2, validating population) of 205 isolates containing an additional 80 *M. tuberculosis* isolates from 1998 and 80 isolates from 2004, both collected in England and Wales (two different time windows were chosen to ensure that the results seen were reproducible and not unique to 1998, the year in which the panel 1 strains were isolated), and 45 *M. bovis* isolates from a panel previously described (14), were analyzed using MIRU-VNTR loci. The *M. tu-*

* Corresponding author. Mailing address: HPA-MRU, Clinical Research Centre, Barts & The London, Queen Mary's School of Medicine and Dentistry, 2 Newark St., Whitechapel, London, E1 2AT, United Kingdom. Phone: 44 207 377 5895. Fax: 44 207 5393 459. E-mail: francis.drobniewski@qmul.ac.uk.

† Present address: University Hospital Lewisham, Lewisham High Street, London, United Kingdom.

rich4/zam-aem/zam-aem/zam01205/zam6156d05z | xppws | S=1 | 10/19/05 | 10:40 | Art: 1165-05

TABLE 1. Primer sequences for the macroarray

Gene	Primers (5′–3′)	Product size (bp)
oxyR	ATCGCCGCCAAGAGGTGCTA TCACGCACTGCACGACGGT	311
katG	TGTCCCGTCGTGGGTCATAT TTGTCCAAGCTGGCGTTGT	370
rpoB	TGCGTGTGTATGTGGCTCAGAAA CGCCGTGGGTGTTCAAAATAAT	159
rpoB	GTAAGGCGCAGTTCGGTGG TTTGAGCAGCACCTTGAACGA	203

berculosis isolates were all randomly chosen from 1998 and 2004. The *M. bovis* strains included all available isolates between 1997 and 2001 collected at the Health Protection Agency National Mycobacterium Reference Unit (HPA-MRU). Panel 2 was characterized with a macroarray developed in this study to define sSNPs seen in each of the five lineages. The validation was performed blinded. Results were compared only once a lineage based on each technique had been determined.

Validation of the macroarray. The macroarray was validated on a subpanel of 46 previously genotyped isolates from panel 1 (1).

MIRU-VNTR analysis. Twelve-MIRU fragment analysis was carried out on a CEQ2000 DNA capillary sequencer (Beckman Coulter, Fullerton, Calif.) as described previously (19) and was supplemented with three loci: ETR-A, ETR-B, and ETR-C (13). Three capillaries were used per isolate. Capillary 1 was used for MIRU-4 and -16 (dye2), MIRU-2 and -24 (dye3), and MIRU-10 and -23 (dye4). Capillary 2 was used for MIRU-39 and ETR-A (dye2), MIRU-27 and ETR-B (dye3), and MIRU-31 and -40 (dye4). Capillary 3 was used for MIRU-20 (dye2), MIRU ETR-C (dye3), and MIRU-26 (dye4). MIRU-VNTR profiles were analyzed using a categorical similarity coefficient (which scores all characters equally and considers any different state as no match and any identical state as a full match) and were displayed using the unweighted-pair group method using average linkages.

DNA macroarray. Biotin-labeled primers, based on the published sequences of *M. tuberculosis* H37Rv and CDC1551, were designed to flank the four sSNPs of interest (in *oxyR*, *katG*, and *rpoB*) (Table 1). PCR was carried out in a final volume of 20 μl, containing 5 μl sterile distilled water, 10 μl 2× reaction buffer (3 mM $MgCl_2$, 3.2 mM deoxynucleoside triphosphates, 2× NH_4 buffer), 5 μl of primer mix (2 μM each primer), 0.2 μl *Taq* polymerase (5 U/μl; Bioline, London, United Kingdom), and 1 μl template DNA. Amplification was performed in 0.2-μl thin-walled 96-well plates (Alpha, Eastleigh, United Kingdom) in a Perkin-Elmer Cetus 9600 Thermocycler with the following program: 3 min at 95°C, followed by 30 cycles of 15 s at 95°C, 30 s at 60°C, and 1 min at 72°C, and one final cycle of 5 min at 72°C.

Two probes were designed per region of interest, one with the wild-type sequence according to the published sequences for *M. tuberculosis* H37Rv and CDC1551, the other containing the sSNP of interest (1). Each probe was 19 to 25 nucleotides long and contained a poly(T) tail of 20 bases (Table 2).

Each oligonucleotide (20 μM containing 0.001% bromophenol blue), an ink dot (for membrane orientation), and a color detection control (2 μM of *oxyR* primer) were dotted onto a nylon membrane (see Fig. 1) and fixed by UV cross-linking. The membrane was washed twice in wash solution (0.5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]–0.1% sodium dodecyl sulfate [SDS]) for 5 min and allowed to air dry.

Twelve microliters of amplified *oxyR*-37 and *rpoB*-3243 from each sample were combined. *katG*-87 and *rpoB*-2646 amplified products were first diluted 1:10, and then 12 μl was added. The DNA mix was denatured at 100°C for 10 min. Hybridization and detection were performed as previously described (24) with the following modifications. Denatured DNA and 500 μl of hybridization solution (5× SSPE [1× SSPE is 0.18 M NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA {pH 7.7}]–0.5% SDS) were incubated with a membrane for 1 h at 60°C. Membranes were washed twice in stringent wash buffer (0.3× SSPE–0.5% SDS) for 1 min at room temperature and then once in stringent wash buffer at 60°C for 10 min. The following washes were all carried out at room temperature. Membranes were washed twice in rinse buffer (0.1 M Tris, 0.1 M NaCl, pH 7.5) and then incubated for 1 min in 5 ml rinse buffer containing 0.5% blocking reagent before addition of streptavidin-alkaline phosphate (Roche, East Sussex, United Kingdom) and incubation for 30 min. Membranes were washed twice with rinse buffer (0.1 M Tris, 0.1 M NaCl, pH 7.5) for 1 min and then equilibrated in a second rinse buffer (0.1 M Tris, 0.1 M NaCl, pH 9.5) before incubation in a color development solution (75-mg/ml nitroblue tetrazolium [U.S. Biochemicals, Cleveland, Ohio] in 70% dimethylformamide and 50-mg/ml 5-bromo-4-chloro-3-indolylphosphate [U.S. Biochemicals] in 100% dimethylformamide) for 1 h.

RESULTS

MIRU-VNTR analysis of panel 1. A test population of 312 *M. tuberculosis* isolates classified by sSNP analysis into four *M. tuberculosis* lineages plus *M. bovis* (1) was typed using 15-locus MIRU-VNTR profiling. MIRU-VNTR profiles were obtained from 309/312 (99.0%) *M. tuberculosis* isolates (the remaining 3 were removed from further analysis) and 4/4 (100%) *M. bovis* isolates. These isolates were examined to determine whether any single MIRU-VNTR polymorphism or combination of polymorphisms was common to each lineage and could potentially be used to define each lineage.

Analysis of MIRU-VNTR profiles showed that the allelic diversity at certain loci was conserved while others varied considerably. No single MIRU-VNTR polymorphism was exclusive to a lineage. By using a combination of polymorphisms, each of the five main phylogenetic branches of the tree could be defined. For example, in lineage I the majority of isolates (95%) contained 3 copies of MIRU-39 and 4 copies of ETR-A, and all isolates contained 4 copies of ETR-C (Table 3). Taken together, 90% of the isolates in lineage I contained all three polymorphisms, i.e., they defined lineage I. The same analysis was performed on the other lineages.

Lineage III was highly conserved, with 98.4% of isolates containing the two polymorphisms defining their group, i.e., 5 copies of MIRU-23 and 2 copies of ETR-C. Lineage IV also displayed a high level of conservation, with 96.7% of isolates containing 2 copies each of MIRU-24 and MIRU-26. The code for lineage II was more variable; therefore, identifying a single defining allele at any locus was difficult. Instead, a combination of alleles at each locus was used—1, 2, or 3 copies of MIRU-16, 2 copies of MIRU-39, and 1 or 2 copies of ETR-B—which defined 92.7% of isolates. *M. tuberculosis* H37Rv, which belongs to lineage II, also displayed a lineage II code.

All *M. bovis* isolates contained the three polymorphisms defining their lineage; however, only four isolates were examined originally by Baker et al. (1). We confirmed that this code was robust by characterizing 45 additional *M. bovis* isolates with the macroarray and MIRU-VNTR (results below).

Macroarray analysis. Based on the sSNP analysis described by Baker et al. (1), a macroarray was developed to identify lineage-defining sSNPs and was validated using 46 strains belonging to panel 1. The macroarray and original sSNP results were concordant, i.e., the macroarray correctly identified the lineage of each isolate based on the detection of probe-defined sSNPs. The hybridization patterns for each lineage are shown in Fig. 1. For example a lineage 1 isolate displays a positive hybridization signal at the wild-type *oxyR*-37, *katG*-87, and *rpoB*-2646 probes, while at *rpoB*-3243 a positive hybridization signal is observed for the mutant probe.

MIRU-VNTR analysis of panel 2. In order to validate the MIRU-VNTR codes defined with the test population, panel 2 (205 isolates) was MIRU-VNTR typed and analyzed for sSNPs with the macroarray. Overall, there was 88.8% concordance between the lineages of all isolates defined by the macroarray and MIRU-VNTR. The *M. bovis* lineage was 88.9% concor-

T4

T5

T6

TABLE 2. Wild-type and mutant probe sequences

Gene and nt position (sSNP)	Wild-type probe			Mutant probe		
	Sequence[a]	Size (bp)		Sequence[a]	Size (bp)	
oxyR-37 (G-A)	CCACCGCGGCGAACGCGCGAAGCCCTTTTTTTTTTTTTTTTTTTTTT	35		GCGGCGAACGCGCGAAACCCGTTTTTTTTTTTTTTTTTTTTTT	35	
katG-87 (C-A)	ACCCCAGTCGAGGGCGGCGGATTTTTTTTTTTTTTTTTTTTTTT	29		ACCCAGTCGAGGGCGGCGGATTTTTTTTTTTTTTTTTTTTTTT	29	
rpoB-2646 (A-C)	CGGCCAGCTTGTCACCGTCGGTTTTTTTTTTTTTTTTTTTTTTT	31		CGGCCAGCTTGTCCCCGTCGGTTTTTTTTTTTTTTTTTTTTTTT	31	
rpoB-3243 (G-A)	CTCCTGCAGGGTGTAGGCGGCATTTTTTTTTTTTTTTTTTTTTTT	31		TCCTGCAGGGTGTAGGCGGCATTTTTTTTTTTTTTTTTTTTTTT	31	

[a] Polymorphic sites are boldfaced.

dant, demonstrating that in the majority of cases *M. bovis* can easily be differentiated from *M. tuberculosis*. Within the *M. tuberculosis* lineages alone, the level of concordance ranged from 81.3% to 100% (Table 4).

Analysis of both panels by MIRU-VNTR. Combining panels 1 and 2 gave 518 isolates (465 *M. tuberculosis* and 49 *M. bovis* isolates) analyzed by both MIRU-VNTR and sSNP (DNA sequencing or the macroarray). Overall, there was 90.9% concordance between lineage-defined MIRU-VNTR phylogenetic codes and sSNP-defined lineages. The percentage of definable isolates ranged from 86.1 to 99.0% for *M. tuberculosis* and was 89.8% for *M. bovis* (Table 5).

Discrepant results. Since sSNPs display neutral variation, they are unlikely to be under any selection pressures. The sSNPs discussed here are concordant with other genetic and phenotypic groupings (including genetic groups 1 to 3, *M. tuberculosis*-specific deletions, spoligotyping, and IS*6110* families) (1), and therefore they are likely to be definitive. Any disagreement with the sSNP-defined lineages was considered discrepant. Overall, 47 (9.1%) discrepant results were observed, i.e., the MIRU-VNTR code did not match the lineage defined by sSNP analysis (Table 6). Of the *M. tuberculosis* isolates, 27/42 (64.3%) did not match any of the defined MIRU-VNTR codes but had a MIRU-VNTR code very similar to that for one of the four lineages, and in each case this matched the lineage defined by sSNP analysis. For example, five isolates with lineage II discrepancy b had a VNTR code similar to that of lineage II. These isolates differed at MIRU-39, possessing 1 copy instead of 2 copies. Three isolates with lineage II discrepancy g contained 5 copies of MIRU-16 and 3 copies of ETR-B as opposed to the predicted profile, 2 or 3 copies of MIRU-16 and 1 or 2 copies of ETR-B. In all lineages, the allelic diversity at the discrepant loci displayed a stepwise trend. For example, isolates with discrepancies b to e in lineage II contained either 1 or 3 copies of MIRU-39 instead of 2 copies, and isolates with discrepancy g or i contained 3 copies of ETR-B instead of 1 or 2 copies. These discrepancies support the presence of sublineages within lineage II.

Twelve *M. tuberculosis* isolates had MIRU-VNTR codes for two lineages. Three isolates defined as lineage II by sSNP analysis had the MIRU-VNTR codes for both lineages II and III, and nine isolates defined as lineage IV had the code for this lineage plus another *M. tuberculosis* lineage code. The phylogenetic code for lineage IV was exclusive and was not observed in any other lineage; therefore, it seems that any discrepant isolate containing the MIRU-VNTR code for lineage IV should be designated a lineage IV strain. However, the MIRU-VNTR code for lineage II was not exclusive, making it difficult to determine to which lineage these isolates belonged. Only 2/245 (1.2%) isolates in lineage II had two codes; therefore, the chances of this occurring are probably low. Interestingly, if the code for lineage II is modified by only counting 2 copies of ETR-B instead of 1 and 2 copies, then the code becomes highly conserved and is not observed in any other lineage. However, the increase in specificity is outweighed by the loss in sensitivity, reducing the proportion of isolates defined from 92.7% to 76.3%.

The remaining three discrepant strains were defined as lineage I according to sSNP analysis, but all had the lineage II code and a partial code for lineage I. All three had a Beijing

rich4/zam-aem/zam-aem/zam01205/zam6156d05z | xppws | S=1 | 10/19/05 | 10:40 | Art: 1165-05

TABLE 3. MIRU-VNTR phylogenetic codes for each lineage based on panel 1

Lineage (*n*)[a]	MIRU-VNTR code[b] (no. of isolates with correct repeat no./total isolates [%])	Combined MIRU-VNTR[c]	
		Code	No. of isolates with correct combined repeat no./total isolates (%)
I (20)	39–3 (19/20 [95])	39–3	18/20 (90)
	A–4 (19/20 [95])	A–4	
	C–4 (20/20 [100])	C–4	
II (168)	16–1,2,3 (164/167 [98.2])	16–1,2,3	153/166 (92.7)
	39–2 (160/167 [95.8])	39–2	
	B–1,2 (162/166 [97.6])	B–1,2	
III (62)	23–5 (61/62 [98.4])	23–5	61/62 (98.4)
	C–2 (62/62 [100])	C–2	
IV (62)	24–2 (61/61 [100])	24–2	59/61 (96.7)
	26–2 (60/62 [98.8])	26–2	
M. bovis (4)	10–2 (4/4 [100])	10–2	4/4 (100)
	40–2 (4/4 [100])	40–2	
	C–5 (4/4 [100])	C–5	

[a] Some loci would not amplify; therefore, two isolates from lineage II and one isolate from lineage IV were excluded from the analyses, leaving a total of 309 *M. tuberculosis* isolates in panel 1.

[b] Each code consists of the MIRU (numbers) or ETR (letters) designation followed by a dash and the number of copies. For example, 16–1,2,3 indicates 1, 2, or 3 copies of MIRU-16; A–4 indicates 4 copies of ETR-A.

[c] Combination of MIRU-VNTR producing the most unambiguous definition of each lineage. The number and percentage of isolates tested that were correctly identified using this combination are given.

TABLE 4. Number of isolates in panel 2 belonging to each lineage according to the macroarray and concordance with MIRU-VNTR results

Lineage defined by macroarray	No. of isolates	Defined MIRU-VNTR loci–repeat no.[a]	Comparison of macroarray with MIRU-VNTR	
			Concordance[b]	No. (%) of discrepant results
I	16	39–3, A–4, C–4	13/16 (81.3)	3 (18.8)
II	79	16–1,2,3, 39–2, B–1,2	68/79 (86.1)	11 (13.9)
III	38	23–5, C–2	38/38 (100)	0 (0.0)
IV	27	24–2, 26–2	23/27 (85.2)	4 (14.8)
M. bovis	45	10–2, 40–2, C–5	40/45 (88.9)	5 (11.1)
Total	205		182 (88.8)	23 (11.2)

[a] Combination of MIRU-VNTR (with repeat number at each locus) producing the most unambiguous definition of each lineage. Numbers stand for MIRU (e.g., 39–3 indicates 3 copies of MIRU-39); letters stand for ETR (e.g., A–4 indicates 4 copies of ETR-A).

[b] Expressed as the number of isolates with concordant results/total number of isolates (percent).

spoligotype (spacers 1 to 34 absent, spacers 35 to 43 present) and may represent evolutionary intermediates.

Of the discrepant *M. bovis* isolates, all five had different codes. Again, a stepwise trend was seen among the discrepant isolates. For example, 2, 3, 4, and 5 copies of ETR-C were observed.

Indeterminate isolates. Four isolates not assigned a lineage in the Baker et al. panel (1) because of incomplete MLST data were analyzed by MIRU-VNTR to determine whether they could be classified using these codes. Three isolates had the MIRU-VNTR code for lineage II and none of the other codes. The remaining isolate had a code similar to that for lineage I except that it had 5 copies of ETR-A instead of 4. sSNP macroarray analysis confirmed that three of the isolates were indeed from lineage II and the remaining isolate was from lineage I.

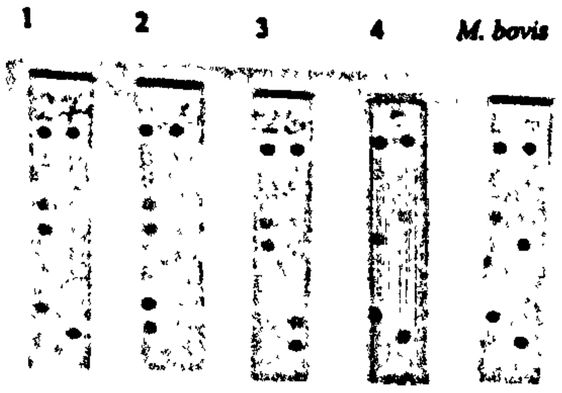


FIG. 1. Hybridization patterns corresponding to the sSNP defining each lineage. Orientation of probes from left to right and top to bottom: ink spot, color development control, four blanks, wild-type *oxyR*-37, mutant *oxyR*-37, wild-type *katG*-87, mutant *katG*-87, four blanks, wild-type *rpoB*-2646, mutant *rpoB*-2646, wild-type *rpoB*-3243, and mutant *rpoB*-3243. Blanks were included to extend the length of the macroarray for ease of handling.

MIRU-VNTR discrimination. Once a population of strains has been assigned to a lineage, it would be useful for epidemiological studies to be able to subdivide lineages into epidemiologically linked and unique isolates. The discriminatory power of each MIRU-VNTR locus was examined, and a specific set of MIRU-VNTR capable of subdividing each lineage was determined (Table 7). Because the MIRU-VNTR used to define each lineage were conserved, they would not be useful for epidemiological studies, where the ability to detect maximum diversity is required. A combination of six loci (MIRU-16, -26, -27, -31, and -20 and ETR-A), when applied to lineage I isolates, was able to generate the same degree of discrimination as the 15-locus MIRU-VNTR overall. The number of MIRU-VNTR loci could not be reduced for lineage II without reducing the discrimination. Nine loci in lineage III (MIRU-10, -16, -20, -26, -27, -31, -39, and -40 and ETR-B) and eight in lineage IV (MIRU-4, -10, -31, -39, -40, and -23, ETR-A, and ETR-B) differentiated all isolates. The greatest discrimination within the *M. bovis* lineage was achieved with eight loci: MIRU-23, -24, -26, and -27, ETR-A, ETR-B, and a combination of two MIRU-VNTR from MIRU-4, -31, and -39. However, if a lower level of discrimination is acceptable, then an even smaller number of MIRU-VNTR can be used, which may be of value for laboratories with resource constraints. For

TABLE 5. Number of isolates belonging to each lineage according to sSNP analysis and concordance with MIRU-VNTR results for all *M. tuberculosis* and *M. bovis* isolates tested

Lineage defined by sSNP	Total no. of isolates tested[a]	Concordance with MIRU-VNTR results (%)
I	36	31 (86.1)
II	245	220 (89.8)
III	100	99 (99.0)
IV	88	77 (87.5)
M. bovis	49	44 (89.8)
Total	518	471 (90.9)

[a] Excludes three isolates from panel 1 that did not amplify.

rich4/zam-aem/zam-aem/zam01205/zam6156d05z | xppws | S=1 | 10/19/05 | 10:40 | Art: 1165-05

TABLE 6. Discrepancies observed in each lineage from a total population of 518 isolates

Lineage	Defined MIRU-VNTR loci–repeat no.[a]	Discrepancy	Discrepant MIRU-VNTR code(s)[b]	No. of times observed
I	39–3, A–4, C–4	a	**39–2**, A–**3**, C–4 + code for lineage II	1
		b	**39–2**, A–4, C–4 + code for lineage II	2
		c	39–3, A–**3**, C–4	1
		e	**39–4**, A–4, C–4	1
II	16–1,2,3, 39–2, B–1,2	a	Codes for lineages II and III present	3
		b	16–3, **39–1**, B–2	5
		c	16–3, **39–1**, B–1	2
		d	16–1, **39–3**, B–1	1
		e	16–3, **39–3**, B–2	4
		f	**16–4**, 39–2, B–2	4
		g	**16–5**, 39–2, **B–3**	3
		h	16–1, 39–2, **B–3**	2
		i	16–3, 39–2, **B–3**	1
III	23–5, C–2	a	**23–3**, C–2	1
IV	24–2, 26–2	a	Codes for lineages I and IV present	6
		b	Codes for lineages II and IV present	2
		c	Codes for lineages III and IV present	1
		d	24–2, **26–1**	2
M. bovis	10–2, 40–2, C–5	a	10–2, 40–2, **C–2**	1
		b	10–2, 40–2, **C–3**	1
		c	10–2, 40–2, **C–4**	1
		d	**10–6**, 40–2, **C–4**	1
		e	10–2, **40–3**, C–5	1

[a] Numbers stand for MIRU (e.g., 16–1,2,3 indicates that there are 1, 2, or 3 copies of MIRU-16); letters stand for ETR (e.g., A–4 indicates that there are 4 copies of ETR-A).
[b] Discrepant loci are boldfaced.

example, by allowing 4 or fewer discrepancies per lineage (i.e., those isolates that are clustered by using a smaller panel of MIRU-VNTR compared to the clustering seen with all 15 MIRU-VNTR), 4, 12, 7, 7, or 6 MIRU-VNTR would be sufficient to produce the same level of intralineage discrimination as the whole panel when applied to lineage I, II, III, or IV or *M. bovis*, respectively (Table 7).

DISCUSSION

The identification of key phylogenetic markers within populations of isolates is important in order to aid in our understanding of *M. tuberculosis* complex evolution and pathogenesis. A series of such markers in *M. tuberculosis* and *M. bovis* have been described previously by Baker et al. (1), who defined

TABLE 7. MIRU-VNTR loci used to subdivide each lineage into true clustered and true unique isolates

Lineage (*n*)	No. of clusters (range)	No. of clustered isolates	No. of unique isolates	MIRU-VNTR (no. of discrepancies)[a]	Additional MIRU-VNTR for total discrimination (no. of discrepancies)[b]
I (36)	6 (2–9)	27	9	16, 26, 27, 31 (4)	A (1) A, 20 (0)
II (245)	30 (2–13)	120	125	4, 10, 16, 20, 23, 26, 27, 31, 39, 40, A, C (4)	B (2) B, 2 (1) B, 2, 24 (0)
III (100)	15 (2–13)	63	37	10, 16, 26, 31, 39, 40, B (2)	20 (1) 20, 27 (0)
IV (88)	7 (2–4)	19	69	4, 10, 31, 39, A, B, 40 (2)	23 (0)
M. bovis (49)	5 (2 only)	10	39	23, 24, 26, 27, A, B (3)	Combination of two MIRU from 4, 31, and 39 (0)

[a] Numbers indicate MIRU (e.g., 16 stands for MIRU-16); letters indicate ETR (e.g., A stands for ETR-A). Isolates could initially be screened with a smaller panel of MIRU-VNTR in a resource-constrained laboratory wishing to produce a high level of intralineage discrimination for epidemiological purposes at a lower cost. This column shows the maximum number of MIRU-VNTR loci required to produce four or fewer discrepancies (i.e., isolates which are clustered by using a smaller panel of MIRU-VNTRs compared to clustering with all 15 MIRU-VNTR).
[b] Additional MIRU-VNTR loci that would need to be added to the smaller panel to give the same level of intralineage discrimination of clustered isolates as the whole panel.

rich4/zam-aem/zam-aem/zam01205/zam6156d05z | xppws | S=1 | 10/19/05 | 10:40 | Art: 1165-05

five distinct lineages. MIRU-VNTR is a rapid, high-throughput typing tool that provides useful epidemiological markers for use in tuberculosis outbreaks and control programs. If the numeric codes by which MIRU-VNTR types are expressed can also define phylogenetic lineages, very large populations of *M. tuberculosis* complex isolates can be used in *M. tuberculosis* evolution studies, since this typing system is beginning to be used routinely in some centers both in the United Kingdom and globally.

In this study MIRU-VNTR was successfully applied to a test population to determine phylogenetic codes capable of defining each of the lineages described previously. To validate these MIRU-VNTR codes, an additional panel was analyzed. This panel was characterized with a portable sSNP-based macroarray. Developing a macroarray to identify known sSNPs in a population allowed rapid identification to species level and identification of key lineages. The macroarray eliminated the need for time-consuming, more-costly sequencing and so proved useful as a preliminary screening tool for *M. tuberculosis* and *M. bovis* populations.

Overall, the MIRU-VNTR phylogenetic codes were able to define 90.9% of sSNP-characterized *M. tuberculosis* and *M. bovis* isolates by using a combination of just 10 of the 15 loci examined.

Lineage I accommodates the Beijing family strains. In this study, 86.1% of the Beijing isolates were defined using the MIRU-VNTR lineage I code. Although MIRU-VNTR codes are not as definitive as spoligotyping, *dnaA-dnaN* (7, 12, 18), and sSNP analysis, the data shown here demonstrate that it is a good additional screening marker for Beijing-type strains. Ferdinand et al. (11) genotyped spoligotype-defined families using the 12-MIRU system. They found that all Beijing isolates could be characterized using a maximum of six MIRU; since the focus was to characterize all isolates, the loci and repeat numbers used were not the same for all Beijing isolates. A single code for the identification of Beijing isolates may be more useful before, or in parallel with, discrimination between these isolates.

Lineage III contained the Central Asian family (CAS)/Delhi family of strains (2, 12). The MIRU-VNTR code for this family, 5 copies of MIRU-23 and 2 copies of ETR-C, defined 99% of the isolates, making the differentiation of CAS/Delhi strains within a population simple and specific.

Locus MIRU-24 has previously been used to classify *M. tuberculosis* strains into two groups: those containing 1 copy and those containing 2 or 3 copies (11). Our MIRU-VNTR data are in agreement with those of Ferdinand et al. (11), who found that 97.8% of East African-Indian (EAI) strains contained 2 copies of MIRU-24. The EAI strains in our study were found in lineage IV. However, other strains, including *M. bovis* strains, contained 2 copies of MIRU-24; therefore, this locus could not be used to define lineage IV alone. The addition of MIRU-26 resulted in this code becoming exclusive, thus providing a simple code to define this family.

An important observation was that the *M. bovis* phylogenetic code was unique and was not observed for any *M. tuberculosis* isolates. Therefore, this phylogenetic code is an excellent marker for differentiating between these two species.

The stepwise trend seen among discrepant isolates suggests that evolution is bidirectional in that the number of repeat copies either increases or decreases over time. Indeed, certain loci show more allelic diversity than others, which may be due to different selection pressures acting at different loci. The molecular clocks of some loci may be faster than others, and certain loci may tolerate more polymorphisms than others. The exact function of MIRU is not known, and it is not known whether individual MIRU have different functions. Their high variability may be a way for the bacteria to adapt to a new environment. Changes in the copy number of tandem repeats may affect gene regulation (29) and have been shown to affect the expression of genes involved in adaptive responses (22). In *Haemophilus influenzae*, variability of repeat number has an effect on virulence (22, 31). These findings suggest that a similar phenomenon could be occurring in the MIRU-VNTR of *M. tuberculosis* complex.

MIRU-VNTR can simultaneously define phylogeny and differentiate strains, as demonstrated in this study. A lineage-specific panel of MIRU-VNTR was used subsequently to subdivide each lineage into clustered and unique strains, thus eliminating the need to analyze all 15 MIRU for every isolate. The advantage of using a smaller panel of MIRU-VNTR is a reduction in the cost of the analysis. In molecular epidemiological studies, total discrimination would normally be required; therefore, isolates could be screened initially with a small MIRU-VNTR panel, followed by analysis of clustered isolates with additional MIRU-VNTR for total discrimination (with additional typing techniques as required).

MIRU-VNTR can characterize a population of *M. tuberculosis* and *M. bovis* isolates for phylogenetic/evolutionary purposes. It is also suggested that MIRU-VNTR is able to provide information on sublineages and molecular clocks. MIRU-VNTR analysis is a high-throughput, reproducible method, producing digital results that are readily portable between laboratories, and so is suited to phylogenetic studies as well as having a high discriminatory value for molecular epidemiological investigations. Conversely, although genotyping by indexing genetically neutral variation provides a robust phylogeny for *M. tuberculosis*, its value is limited for definitive epidemiological typing. This study demonstrated that MIRU-VNTR profiling can be used for rapid classification of *M. tuberculosis* and *M. bovis* isolates. Future studies should investigate the potential for extending the use of MIRU-VNTR to identify the other members of the *M. tuberculosis* complex: *Mycobacterium africanum*, *Mycobacterium microti*, and *Mycobacterium canetti*.

ACKNOWLEDGMENTS

The study was funded in part by the United Kingdom Department of Health, Department For International Development (CNTR 00 00134), and the Health Protection Agency.

We thank all our colleagues at the MRU for their help and support. We particularly thank Krishna Gopaul for invaluable advice.

REFERENCES

1. **Baker, L., T. Brown, M. C. Maiden, and F. Drobniewski.** 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg. Infect. Dis. **10**:1568–1577.
2. **Bhanu, N. V., D. van Soolingen, J. D. van Embden, L. Dar, R. M. Pandey, and P. Seth.** 2002. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. Tuberculosis (Edinburgh) **82**:105–112.
3. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **99**:3684–3689.

rich4/zam-aem/zam-aem/zam01205/zam6156d05z | xppws | S=1 | 10/19/05 | 10:40 | Art: 1165-05

4. **Cowan, L. S., L. Mosher, L. Diem, J. P. Massey, and J. T. Crawford.** 2002. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. J. Clin. Microbiol. 40:1592–1602.
5. **Dale, J. W., H. Al Ghusein, S. Al Hashmi, P. Butcher, A. L. Dickens, F. Drobniewski, K. J. Forbes, S. H. Gillespie, D. Lamprecht, T. D. McHugh, R. Pitman, N. Rastogi, A. T. Smith, C. Sola, and H. Yesilkaya.** 2003. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS*6110*. J. Bacteriol. 185:2555–2562.
6. **Dingle, K. E., F. M. Colles, D. R. Wareing, R. Ure, A. J. Fox, F. E. Bolton, H. J. Bootsma, R. J. Willems, R. Urwin, and M. C. Maiden.** 2001. Multilocus sequence typing system for *Campylobacter jejuni*. J. Clin. Microbiol. 39:14–23.
7. **Drobniewski, F., Y. Balabanova, M. Ruddy, L. Weldon, K. Jeltkova, T. Brown, N. Malomanova, E. Elizarova, A. Melentyev, E. Mutovkin, S. Zhakharova, and I. Fedorin.** 2002. Rifampin- and multidrug-resistant tuberculosis in Russian civilians and prison inmates: dominance of the Beijing strain family. Emerg. Infect. Dis. 8:1320–1326.
8. **Dye, C., G. P. Garnett, K. Sleeman, and B. G. Williams.** 1998. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. Lancet 352:1886–1891.
9. **Enright, M. C., and B. G. Spratt.** 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology 144:3049–3060.
10. **Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes.** 1998. IS*6110* transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. J. Bacteriol. 180:2102–2109.
11. **Ferdinand, S., G. Valetudie, C. Sola, and N. Rastogi.** 2004. Data mining of *Mycobacterium tuberculosis* complex genotyping results using mycobacterial interspersed repetitive units validates the clonal structure of spoligotyping-defined families. Res. Microbiol. 155:647–654.
12. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, G. Kallenius, E. Kassa-Kelembho, T. Koivula, H. M. Ly, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi.** 2002. Global distribution of *Mycobacterium tuberculosis* spoligotypes. Emerg. Infect. Dis. 8:1347–1349.
13. **Frothingham, R., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 144:1189–1196.
14. **Gibson, A. L., G. Hewinson, T. Goodchild, B. Watt, A. Story, J. Inwald, and F. A. Drobniewski.** 2004. Molecular epidemiology of disease due to *Mycobacterium bovis* in humans in the United Kingdom. J. Clin. Microbiol. 42:431–434.
15. **Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser.** 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics 162:1533–1543.
16. **Hawkey, P. M., E. G. Smith, J. T. Evans, P. Monk, G. Bryan, H. H. Mohamed, M. Bardhan, and R. N. Pugh.** 2003. Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS*6110*-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. J. Clin. Microbiol. 41:3514–3520.
17. **Kam, K. M., C. W. Yip, L. W. Tse, K. L. Wong, T. K. Lam, K. Kremer, B. K. Au, and D. van Soolingen.** 2005. Utility of mycobacterial interspersed repetitive unit typing for differentiating multidrug-resistant *Mycobacterium tuberculosis* isolates of the Beijing family. J. Clin. Microbiol. 43:306–313.
18. **Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Soolingen, J. M. Musser, and B. N. Kreiswirth.** 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non-random integration in the *dnaA-dnaN* region. Tuber. Lung Dis. 79:31–42.
19. **Kwara, A., R. Schiro, L. S. Cowan, N. E. Hyslop, M. F. Wiser, H. S. Roahen, P. Kissinger, L. Diem, and J. T. Crawford.** 2003. Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates. J. Clin. Microbiol. 41:2683–2685.
20. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA 95:3140–3145.
21. **Mokrousov, I., O. Narvskaya, E. Limeschenko, A. Vyazovaya, T. Otten, and B. Vyshnevskiy.** 2004. Analysis of the allelic diversity of the mycobacterial interspersed repetitive units in *Mycobacterium tuberculosis* strains of the Beijing family: practical implications and evolutionary considerations. J. Clin. Microbiol. 42:2438–2444.
22. **Moxon, E. R., P. B. Rainey, M. A. Nowak, and R. E. Lenski.** 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr. Biol. 4:24–33.
23. **Musser, J. M., A. Amin, and S. Ramaswamy.** 2000. Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. Genetics 155:7–16.
24. **Nikolayevsky, V., T. Brown, Y. Balabanova, M. Ruddy, I. Fedorin, and F. Drobniewski.** 2004. Detection of mutations associated with isoniazid and rifampin resistance in *Mycobacterium tuberculosis* isolates from Samara Region, Russian Federation. J. Clin. Microbiol. 42:4498–4502.
25. **Raviglione, M. C., D. E. Snider, Jr., and A. Kochi.** 1995. Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic. JAMA 273:220–226.
26. **Roring, S., A. N. Scott, H. R. Glyn, S. D. Neill, and R. A. Skuce.** 2004. Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. Vet. Microbiol. 101:65–73.
27. **Sola, C., I. Filliol, E. Legrand, I. Mokrousov, and N. Rastogi.** 2001. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS*1081*, IS*6110*, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J. Mol. Evol. 53:680–689.
28. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA 94:9869–9874.
29. **Supply, P., J. Magdalena, S. Himpens, and C. Locht.** 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol. Microbiol. 26:991–1003.
30. **Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht.** 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J. Clin. Microbiol. 39:3563–3571.
31. **van Belkum, A., S. Scherer, W. van Leeuwen, D. Willemse, L. van Alphen, and H. Verbrugh.** 1997. Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. Infect. Immun. 65:5017–5027.